



UNIVERSIDADE DE ÉVORA

# OntoSPARES: da linguagem natural às ontologias

Contributos para a classificação automática de dados históricos (séc. XVI-XVIII)

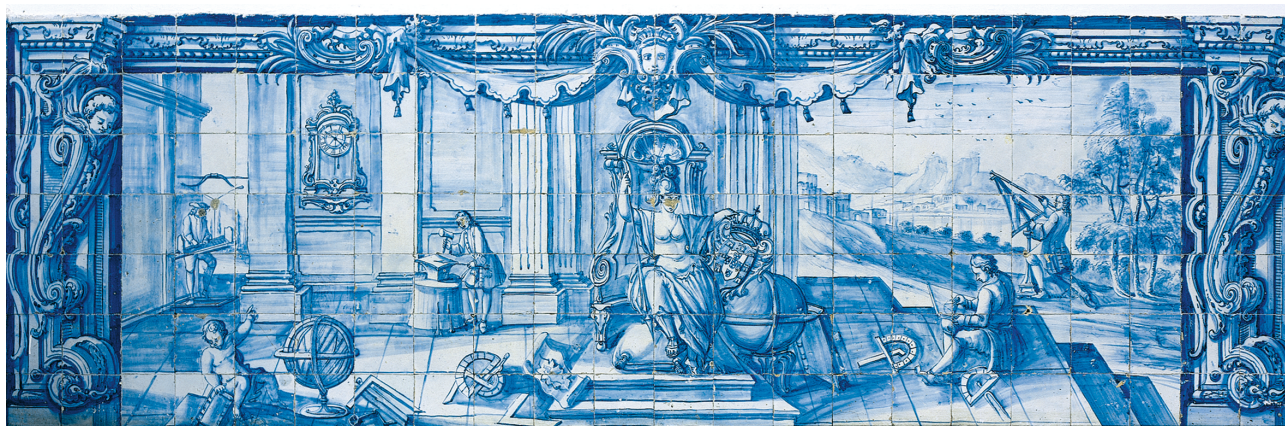
---

*Albertina Maria Gomes Ferreira*

Tese apresentada à Universidade de Évora  
para obtenção do Grau de Doutor em Informática

Orientadores: *Carlos Pampulim Caldeira*  
*Fernanda Olival*

Évora, novembro de 2016



INSTITUTO DE INVESTIGAÇÃO E FORMAÇÃO AVANÇADA





Departamento de Informática

Albertina Maria Gomes Ferreira

Orientação *Carlos Pampulim Caldeira*  
*Fernanda Olival*

Informática

Tese

Évora, novembro de 2016





*Ao Vasco, à Rebeca e à Salomé,  
que muitas vezes não compreendem o que faço  
mas que estão sempre presentes.*



# Agradecimentos

Ao finalizar este trabalho agradeço a todos aqueles que fizeram parte deste percurso, que me apoiaram e que acreditaram que chegaria ao fim.

O meu primeiro agradecimento é dirigido ao Professor Carlos Caldeira. A ele agradeço a orientação científica do trabalho, os comentários, as críticas, a partilha de ideias, o apoio, num trajeto que nem sempre foi linear.

Os resultados obtidos nesta dissertação são o resultado de muitas sessões de trabalho numa área de conhecimento sobre a qual pouco sabia. A integração numa equipa com um projeto já em curso foi decisiva no desenvolvimento do trabalho. Foram também fundamentais neste percurso os recursos disponibilizados (acesso à base de dados SPARES). Assim, agradeço à Professora Fernanda Olival a coorientação deste trabalho, as suas palavras de incentivo, a forma como sempre me recebeu, o entusiasmo que sempre manifestou pelos resultados obtidos.

Aos Professores Irene Rodrigues e Paulo Quaresma agradeço a disponibilidade demonstrada e os conhecimentos transmitidos em áreas fundamentais no decurso do trabalho.

Agradeço ao Professor Francisco Coelho pela adequação do modelo Latex proposto pelo IIFA às especificidades do presente trabalho.

Agradeço ainda o apoio institucional:

- da Escola Superior Agrária de Santarém, que sempre apoiou os meus projetos académicos;
- da Escola de Ciência e Tecnologia da Universidade de Évora, pelo acolhimento no curso de doutoramento em Informática;
- da Fundação para a Ciência e Tecnologia e do CIDEHUS (Centro Interdisciplinar de História, Culturas e Sociedades), pela integração no projeto de investigação PTDC/HIS-HIS/118227/2010 – Grupos intermédios em Portugal e no Império Português: as familiaturas do Santo Ofício (c. 1570-1773).

Ao Manel, à Ana e à Cláudia agradeço a cooperação sem limites, a amizade, a preocupação e o apoio que sempre demonstraram.

Aos amigos que, de uma forma ou de outra, me acompanharam ao longo destes anos, os meus sinceros agradecimentos.

À minha irmã Fátima um agradecimento especial por estar sempre disponível para rever os meus “manuscritos”.

Um enorme agradecimento à minha amiga Anabela, que percorreu ao meu lado este caminho. Pelas horas de trabalho, pela presença, pelas gargalhadas, mas também pelas lágrimas. Contigo tudo foi mais fácil ...

Por último, e não menos importante, à minha família, por todo o carinho e constante apoio, por me tolerarem nos momentos mais difíceis e ... desculparem as minhas ausências.

**A todos o meu MUITO OBRIGADA!**

# Conteúdo

Conteúdo	ix
Lista de Figuras	xiii
Lista de Tabelas	xv
Lista de Siglas	xvii
Sumário	xix
Abstract	xxi
1 Introdução	1
2 Processamento de linguagem natural	5
2.1 Níveis de processamento de linguagem natural . . . . .	6
2.1.1 Segmentação de texto . . . . .	7
2.1.2 Análise lexical . . . . .	7
2.1.3 Análise sintática . . . . .	8
2.1.4 Análise semântica . . . . .	8
2.1.5 Análise pragmática . . . . .	9
3 Extração automática	11
3.1 Extração automática de informação textual . . . . .	12
3.1.1 Áreas de aplicação . . . . .	13
3.2 Métricas de avaliação . . . . .	15

<b>4</b>	<b>Ontologias</b>	<b>17</b>
4.1	Definir . . . . .	17
4.2	Classificar . . . . .	19
4.3	Construir . . . . .	20
4.4	Linguagens . . . . .	21
<b>5</b>	<b>Recursos utilizados e metodologias aplicadas</b>	<b>23</b>
5.1	Ferramentas de desenvolvimento . . . . .	23
5.1.1	Protégé . . . . .	24
5.2	Cenário de referência . . . . .	25
5.3	Identificação de palavras-chave . . . . .	31
5.3.1	Levantamento de valores únicos . . . . .	31
5.3.2	Triagem de palavras-chave . . . . .	31
5.4	Extração de eventos . . . . .	34
5.5	Remoção de <i>stopwords</i> e aplicação de <i>stemming</i> . . . . .	38
5.6	Classificação de eventos ... . . . .	40
5.6.1	... a partir das palavras-chave obtidas manualmente . . . . .	40
5.6.2	... a partir das palavras-chave obtidas por remoção de <i>stowords</i> e aplicação de <i>stemming</i> . . . . .	41
5.7	Criação de cenários ontológicos . . . . .	42
<b>6</b>	<b>Desenvolvimento das ontologias</b>	<b>47</b>
6.1	Cenário I: transposição do modelo SPARES para ontologia . . . . .	47
6.1.1	Classes . . . . .	47
6.1.2	Propriedades de tipo de dados . . . . .	48
6.1.3	Propriedades de objeto . . . . .	49
6.2	Cenário II: criação de nova ontologia . . . . .	49
6.2.1	Classes . . . . .	49
6.2.2	Propriedades de tipo de dados . . . . .	52
6.2.3	Propriedades de objeto . . . . .	52
6.3	Inserção de instâncias . . . . .	56
6.4	Interrogar a ontologia . . . . .	57
<b>7</b>	<b>Discussão de resultados</b>	<b>61</b>
<b>8</b>	<b>Conclusões</b>	<b>63</b>
8.1	Desafios futuros . . . . .	64

<i>CONTEÚDO</i>	xi
A Publicações consultadas: ontologia e/ou Protégé	67
B Palavras-chave obtidas manualmente	77
C Lista de stopwords	99
D Eventos sem stopwords, com aplicação de stemming	103
E Protótipo em desenvolvimento	109
Referências bibliográficas	113





# Lista de Figuras

2.1	Etapas na análise de PLN (Adaptado) (Dale, 2010).	7
5.1	Agrupamento de ocupações (designação da época) por setores de atividade.	28
5.2	Distribuição dos estatutos por setor de atividade.	30
5.3	Distribuição dos estatutos por áreas de ocupação na designação da época.	30
5.4	Distribuição dos estatutos por setor de ocupação, após triagem de palavras-chaves.	34
5.5	Apresentação parcial de uma interface.	35
5.6	Interface para preenchimento de eventos.	35
5.7	Exemplo de evento preenchido na tabela Comissarios.	35
5.8	Pesquisa que aplica as substituições referidas na Tabela 5.11.	37
5.9	Pesquisa que aplica as substituições referidas na Tabela 5.12.	37
5.10	Exemplo das ocupações após execução das pesquisas apresentadas nas figuras 5.8 e 5.9.	37
5.11	Pesquisa que transforma os eventos de ocupação em texto que será processado por ferramentas de PLN.	37
5.12	Exemplo das ocupações após execução da pesquisa apresentada na Figura 5.11.	37
5.13	Criação da tabela Evento.	38
5.14	Criação e preenchimento do campo TextoVetor.	40
5.15	Pesquisas utilizadas para a classificação dos eventos.	41
5.16	Instruções utilizadas para a atribuição de ocupação, atividade e estatuto aos eventos cujas palavras-chave foram obtidas automaticamente.	42
5.17	<i>Mappings</i> construídos para povoar a classe Localizacoes.	43
5.18	SPARQL <i>Query</i> : instruções iniciais.	44
5.19	Prefixo declarado para OntoSPARES.	44
5.20	Sintaxe de uma consulta em SPARQL (Fonte: W3C (2008)).	44

6.1	Classes desenhadas para o cenário I. . . . .	48
6.2	Classes e propriedades de tipo de dados obtidas através da ferramenta <i>Generate ontology and mappings</i> . . . . .	48
6.3	Classes e respetivas propriedades de tipo de dados. . . . .	49
6.4	Classes principais da ontologia - cenário II. . . . .	51
6.5	Subclasses da classe EstatutoEpoca. . . . .	51
6.6	Subclasses da classe SetorAtividade. . . . .	51
6.7	Subclasses da classe Ocupacao. . . . .	52
6.8	Propriedades de objeto inversas. . . . .	56
6.9	Instância: IDministro=31187. . . . .	56
6.10	Instância: Localizacao-1013. . . . .	56
6.11	Resultado parcial da execução da consulta 1. . . . .	58
6.12	Resultado parcial da execução da consulta 2. . . . .	58
6.13	Resultado da execução da consulta 3. . . . .	58
6.14	Resultado parcial da execução da consulta 4. . . . .	58
6.15	Resultado parcial da execução da consulta 5. . . . .	58
6.16	Resultado da execução da consulta 6. . . . .	59
6.17	Resultado parcial da execução da consulta 7. . . . .	59
6.18	Resultado da execução da consulta 8. . . . .	59
6.19	Resultado da execução da consulta 9. . . . .	59
6.20	Resultado da execução da consulta 10. . . . .	59
E.1	Distribuição dos estatutos para a província da Beira. . . . .	110
E.2	Contabilização dos setores de atividade para a província da Estremadura e ocupação Artes Me- cânicas. . . . .	110
E.3	Distribuição, por província, do ofício “Vive de sua fazenda”. . . . .	111

# Lista de Tabelas

3.1	Formulação das métricas de avaliação. . . . .	15
3.2	Publicações vs métricas de avaliação. . . . .	16
4.1	Tipos de ontologias (Adaptado) (Guarino, 1998; Guarino et al., 2009). . . . .	19
4.2	Terminologia da SPARQL (Adaptado) (W3C, 2008). . . . .	22
5.1	Editores de ontologias. . . . .	24
5.2	Descrição das tabelas principais da base de dados SPARES. . . . .	25
5.3	Exemplo parcial de um registo retirado do SPARES. . . . .	26
5.4	Descrição de eventos. . . . .	26
5.5	Designação dos eventos por tipo. . . . .	27
5.6	Tipo de estatuto por ocupação/setor de atividade. . . . .	29
5.7	Exemplo de triagem para a palavra Administrador. . . . .	31
5.8	Palavras-chave com classificação a depender do contexto. . . . .	33
5.9	Exemplos de novos ofícios. . . . .	34
5.10	Ocorrências de XML e carácter de substituição. . . . .	36
5.11	Ocorrências de XML e respetiva substituição. . . . .	36
5.12	Ocorrências de XML substituídas por valor NULL. . . . .	36
5.13	Visualização de evento sem <i>stopwords</i> e após <i>stemming</i> . . . . .	38
5.14	Localização das listas de <i>stopwords</i> . . . . .	39
5.15	<i>Stopwords</i> removidas da lista inicial. . . . .	39
5.16	Palavras introduzidas na lista de <i>stopwords</i> . . . . .	39
5.17	Nomes de freguesias, concelhos ou comarcas irrelevantes para a classificação. . . . .	39
5.18	<i>Stopwords</i> resultantes dos nomes de freguesias, concelhos ou comarcas. . . . .	40

5.19	Exemplo de eventos com a mesma correspondência no campo TextoVetor. . . . .	40
5.20	Exemplo da classificação para 4 eventos. . . . .	41
5.21	Consultas em SPARQL: tipo de resultados obtidos (Adaptado) (W3C, 2008). . . . .	44
6.1	Propriedades de tipo de dados, no cenário I. . . . .	50
6.2	Propriedades de objeto, no cenário I. . . . .	50
6.3	Propriedades de tipo de dados com domínio na classe Ocupacao. . . . .	52
6.4	Propriedades de subclasses de SetorAtividade para subclasses de Ocupação. . . . .	53
6.5	Propriedades de Subclasses de Ocupacao para Subclasses de EstatutoEpoca. . . . .	54
6.6	Propriedades de subclasses de SetorAtividade para subclasses de EstatutoEpoca. . . . .	55
6.7	Consultas realizadas sobre a OntoSPARES. . . . .	57
7.1	Métricas de avaliação aplicadas à classificação com palavras-chave manuais. . . . .	61
7.2	Métricas de avaliação aplicadas à classificação com palavras-chave manuais, após remoção de <i>stopwords</i> e aplicação de <i>stemming</i> . . . . .	62

# Lista de Siglas

CIDEHUS	Centro Interdisciplinar de História, Culturas e Sociedades
DOM	<i>Document Object Model</i>
FCT	Fundação para a Ciência e Tecnologia
HISCO	<i>Historical International Standard Classification of Occupation</i>
ISCO	<i>International Standard Classification of Occupation</i>
ODBA	<i>Ontology Based Data Access</i>
OWL	<i>Web Ontology Language</i>
PLN	Processamento de Linguagem Natural
RDF	<i>Resource Description Framework</i>
SGBDR	Sistema de Gestão de Bases de Dados Relacionais
SO	Santo Ofício
SPARES	Sistema Prosopográfico de Análise de Relações e Eventos Sociais
SPARQL	<i>Simple Protocol and RDF Query Language</i>
SQGM	<i>SPARQL query graph model</i>
SQL	<i>Structured Query Language</i>
XML	<i>Extensible Markup Language</i>
W3C	<i>World Wide Web Consortium</i>
WEB	<i>World Wide Web</i>



# Sumário

O processamento de linguagem natural e as ontologias são ferramentas cuja interação permite uma melhor compreensão dos dados armazenados. Este trabalho, ao associar estas duas áreas aos elementos disponíveis numa base de dados prosopográfica, tornou possível identificar e classificar relacionamentos entre setores de ocupação na forma como eram designados na época, setores de atividade num formato mais próximo do de hoje e o estatuto social que essas incumbências tinham na sociedade coeva. Os dados utilizados são sobretudo de membros do Santo Ofício – do século XVI ao século XVIII. Para atingir este objetivo utilizaram-se algumas descrições textuais de ocorrências da época e outras pouco estruturadas, disponíveis no repositório SPARES. A aplicação de processamento de linguagem natural (remoção de *stopwords* e aplicação de *stemming*), conjugada com a construção de duas ontologias, tornou possível classificar esses dados, permitindo consultas mais eficazes. Ao contribuir para a classificação automática de dados históricos, propõem-se metodologias que podem ser aplicadas em dados de qualquer outra área do conhecimento, especialmente as que lidam com as variáveis de tempo e espaço de forma mais intensa.

**Palavras-chave:** ontologias, linguagem natural, bases de dados históricas, prosopografia





# Abstract

## OntoSPARES: from natural language to ontologies

Contributions to the automatic classification of historical data (16th-18th centuries)

The interaction between the natural language processing and ontologies are tools allowing a better understanding of the data stored. This work, by combining these two areas to the elements available in a prosopographic database, has made possible to identify and classify relationships between occupations of many individuals (in general Holy Office members of the 16th-18th centuries). To achieve this goal the data used was gathered in SPARES repository, including some textual descriptions of the time occurrences. They are all few structured. The application of natural language processing (stopwords removal and stemming application), combined with the construction of two ontologies, made possible to classify those data, allowing a more effective search. By contributing to the automatic classification of historical data, this thesis proposes methodologies that can be applied to data from any other field of knowledge, specially data dealing with time and space variables.

**Keywords:** ontologies, natural language, historical databases, proposography



# 1

## Introdução

A evolução tecnológica permitiu a passagem para suporte digital de informação que estava disponível apenas em suporte analógico. O projeto PTDC/HIS-HIS/118227/2010 - COMPETE/QREN/FEDER - FCOMP-01-0124-FEDER-020722 - INTERGROUPS - Intermediate groups in the Portuguese dominions: the "familiares" of the Holy Office (c. 1570 -1773) permitiu reunir, num único repositório digital, dados que só se encontravam disponíveis em arquivos (designadamente no da Torre do Tombo), com possibilidade de consultas somente presenciais.

Segundo o Centro Interdisciplinar de História, Culturas e Sociedades (CIDEHUS)<sup>1</sup>, este projeto, ao estudar os perfis sociais dos familiares do Santo Ofício (SO) em diferentes contextos, permite analisar os grupos intermédios da sociedade portuguesa da época. Para tal consideraram cinco objetivos principais: "aumentar o conhecimento sobre os grupos intermédios no Antigo Regime português; tentar responder à questão de como se obtinha em Portugal o direito a solicitar uma familiatura, em diferentes etapas cronológicas; conhecer o perfil social e a atuação dos familiares do Santo Ofício, comparando umas zonas do país e do Império com outras e inclusivamente com Espanha".

A criação da base de dados SPARES (Sistema Prosopográfico de Análise de Relações e Eventos Sociais) tornou possível pesquisar a qualquer momento a informação obtida através do projeto acima referenciado, para qualquer finalidade de estudo. Este sistema, com um elevadíssimo número de registos (atingiu os 170.000 em 7 de

---

<sup>1</sup><http://spares.cidehus.uevora.pt/>

agosto de 2016), é continuamente melhorado, permitindo que seja possível dar respostas mais eficientes às questões colocadas por todos aqueles que o utilizam. É um sistema colaborativo, pelo que nele trabalham em simultâneo diversos historiadores, com conhecimentos de Informática que não ultrapassam o patamar do utilizador experiente.

Iremos trabalhar sobre dados históricos que se caracterizam por terem um apego intrínseco às variáveis tempo/espaço. A partir de elementos disponíveis na referida base de dados prosopográfica (de pessoas), optou-se por investir na área das ocupações, por ser um campo de trabalho de elevado interesse para ajudar a classificar o estatuto social dos indivíduos. Para cada personagem importa à partida ter presente pelo menos estes 3 aspetos: 1) setores de ocupação, tal como seriam definidos na época em estudo (séc. XVI-XVIII); 2) setores de atividades mais próximos do seu atual entendimento; 3) estatuto social das ocupações, atribuído na época. Os aspetos mencionados referem-se a diversos indivíduos, sobretudo de membros do SO ou com ele de alguma forma relacionados – do século XVI ao século XVIII.

A área das ocupações tem interesse até para a contemporaneidade. Existe, inclusivamente, um projeto europeu neste domínio, o *International Standard Classification of Occupation* (ISCO), o qual criou suporte para comparações internacionais sobre profissões. O ISCO, contudo, fez pouco para a época aqui tratada.

Pela forma como o SPARES foi preenchido, as técnicas de processamento de linguagem natural (PLN), remoção de *stopwords* e aplicação de *stemming* contribuirão para a classificação dos dados prosopográficos, uma vez que foram utilizadas para o efeito transcrições da época e registos com linguagem pouco controlada. O desafio é melhorar este repositório de dados, torná-lo mais estruturado. Trata-se de uma questão fundamental por mais duas razões, além das aduzidas. Por um lado, por ser um sistema que não teve, intencionalmente, um modelo de dados feito nem em função das fontes, nem do problema de investigação, de modo a servir para um leque alargado de investigadores com interesses na prosopografia; por outro, como sistema colaborativo, há desfazamentos nos registos introduzidos pelos vários participantes, que importa minimizar quando se pretende recuperar informação. A aposta desta tese é desenvolver estratégias que facilitem a classificação dinâmica dos dados e uma mais ajustada recuperação de conhecimentos. Desde logo, qualquer intervenção tem sempre que ter presentes as variáveis tempo e espaço, fundamentais no trabalho do historiador.

A partir desta base de conhecimento, construíram-se duas ontologias, utilizando o *software* Protégé. Estas permitirão realizar pesquisas sobre as classificações obtidas. As ontologias, ao representarem um domínio específico de interesse, podem contribuir para a obtenção de novos conhecimentos. Deste modo, considera-se que a utilização de ontologias, conjugadas com remoção de *stopwords* e aplicação de *stemming*, por comparação a métodos *ad hoc*, são um importante contributo para a classificação de dados prosopográficos.

Este trabalho tem, assim, como objetivos: i) classificar os eventos por setor de ocupação, setor de atividade e estatuto; ii) aplicar aos dados métricas de avaliação da extração de dados; iii) construir ontologias que permitam estruturar a base de conhecimento com uma linguagem pouco controlada. Tudo isto visa por um lado facilitar a classificação das inúmeras ocupações num quadro mais reduzido e inteligível, por outro contribuir para a própria classificação das personagens envolvidas. Assim, pretende-se obter resposta para as seguintes questões:

- A utilização de PLN, nomeadamente o controlo das *stopwords* e aplicação de *stemming*, na classificação automática de eventos é vantajosa relativamente a uma classificação manual num universo com as características que já foram definidas?
- As listas de *stopwords* disponíveis para a Língua Portuguesa serão adequadas ao universo dos dados a classificar?
- Ao interrogar as ontologias criadas é possível gerar novos conhecimentos?
- Os resultados alcançados beneficiarão o desenvolvimento de futuros projetos no CIDEHUS?

Após um enquadramento teórico dos conceitos de PLN (secção 2), extração automática (secção 3) e ontologias (secção 4), far-se-á uma descrição dos recursos que foram utilizados e das metodologias que foram seguidas (secção 5), com a construção de dois cenários distintos. O primeiro cenário resulta de uma transposição do modelo concetual do repositório SPARES. O segundo cenário foi desenhado tendo por base a exploração de dados feita com recurso a dicionários da época, contextos de frase, leque ocupacional dos indivíduos e triagem dos resultados por uma equipa de profissionais. Na secção 6 apresentam-se as ontologias desenvolvidas e na secção 7 tecem-se alguns comentários aos resultados obtidos. Finaliza-se o trabalho com as considerações finais e algumas sugestões de trabalhos futuros nesta área de investigação (secção 8).



# 2

## Processamento de linguagem natural

A língua é a principal manifestação da inteligência humana. Através da linguagem expressam-se desde necessidades básicas a conhecimentos técnicos. Considera-se linguagem natural a que é recorrentemente utilizada no quotidiano para a comunicação humana. Em contraste com as linguagens artificiais, como as de programação, a linguagem natural evolui de geração em geração, e embora tenha regras, estas são pouco estáticas.

Baranow (1983) refere um crescente interesse no PLN, nomeadamente no que diz respeito à implementação de sistemas de diálogo entre o homem e o computador e do incremento da capacidade de memória computacional. Bird, Klein e Loper (2008, para. 1) escreveram no 1º capítulo da sua publicação: *"Today, people from all walks of life - including professionals, students, and the general population - are confronted by unprecedented volumes of information, the vast bulk of which is stored as unstructured text"*. Chisholm (2013) reforça esta ideia ao referir que o volume de dados armazenados e a velocidade com que podem ser consultados são mais elevados a cada dia que passa. No entanto, a capacidade das pessoas para processar e compreender esses dados permanece constante. Para ultrapassar esta limitação humana, o PLN dedica-se a investigar, propor e desenvolver sistemas computacionais que têm a linguagem natural escrita como objeto primário (Grishman, 1986). O PLN pode ser definido num sentido amplo para qualquer tipo de manipulação da linguagem natural em computador. Assim, pode ser algo tão simples como a contagem do número de vezes que a letra t ocorre num parágrafo, ou então pode envolver a "compreensão" de expressões humanas (Bird et al., 2008).

Liddy (2001, para. 2) considera que não existe uma definição única para PLN. Ainda assim oferece a seguinte definição: *"Natural Language Processing is a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications"*. Desta definição, a autora destaca a noção imprecisa de conjunto de técnicas computacionais, pois existem vários métodos ou técnicas que se podem escolher. Refere também que os textos de ocorrência natural podem ser de qualquer idioma, modo ou género. Podem ainda ser orais ou escritos. A única exigência é que eles sejam de uma linguagem utilizada pelo homem para comunicar entre si. Considera ainda que existem vários níveis de análise linguística com diferentes significados e que os sistemas de PLN utilizam diferentes níveis ou combinações desses níveis. O PLN torna, assim, possível que os computadores consigam compreender e interpretar as mais diversas instruções fornecidas por uma determinada linguagem natural.

No contexto do PLN podem-se considerar vários tipos de sistemas, com diferentes níveis de aperfeiçoamento, bem como de desempenho, nomeadamente: dicionários, thesaurus e enciclopédias eletrónicas; sistemas de recuperação de informação; sistemas de extração de informação; sistemas de tradução automática; sistemas de sumarização automática; sistemas de correção ortográfica; sistemas de correção gramatical; analisadores semânticos; analisadores discursivos; sistemas de categorização de textos; sistemas de diálogos; sistemas de auxílio à escrita (Di Felippo & Dias-da Silva, 2009).

Dos sistemas acima referidos, destacam-se os sistemas de extração de informação, tratados mais pormenorizadamente na secção 3, pela sua importância no tratamento automático das línguas naturais.

## 2.1 Níveis de processamento de linguagem natural

O que distingue as aplicações que processam linguagem de outros sistemas de processamento de dados é a sua utilização do conhecimento da língua. Nessas aplicações é importante o sistema poder "recordar" questões anteriores que lhe foram colocadas. Martin e Jurafsky (Martin & Jurafsky, 2000, p. 3) ao colocarem a questão *"How many states were in the United States that year?"*, comentam que se pode equacionar a que ano esta se refere. A resposta seria "ao ano em que Lincoln nasceu". Nestes casos é fundamental o conhecimento de partes anteriores do discurso. Para que questões deste tipo possam ter resposta, são requeridos diversos tipos de conhecimento.

Para Küppers (2013) a informação é baseada em sinais e símbolos e sequências destes. Estas sequências podem ter um sentido ou um significado, podendo a informação dividir-se por três dimensões: i) sintática; ii) semântica; iii) pragmática. A dimensão sintática é entendida como o arranjo ordenado de símbolos bem como das relações entre eles. A dimensão semântica inclui não só as relações entre os símbolos, mas também o seu significado. A dimensão pragmática inclui as relações entre os símbolos, o seu significado, e ainda o efeito destes sobre o destinatário. Este autor considera ainda que as três dimensões constituem uma unidade indissolúvel, não se podendo mencionar uma delas sem se atender às restantes.

Em 2010, Dale advoga que tradicionalmente o PLN considera o processo de análise da linguagem como sendo decomposto numa série de etapas. Para além das três dimensões referenciadas por Küppers (2013) este autor inclui também duas etapas iniciais: segmentação de texto e análise lexical. As frases de um texto são primeiramente analisadas em termos de sintaxe (ordem e estrutura), seguida de uma análise em termos de semântica (significado literal) e, por fim, uma análise pragmática (significado da expressão vocal ou de contexto de texto). Visualizam-se esquematicamente na Figura 2.1 as etapas de análise de PLN apresentadas por Dale (2010).



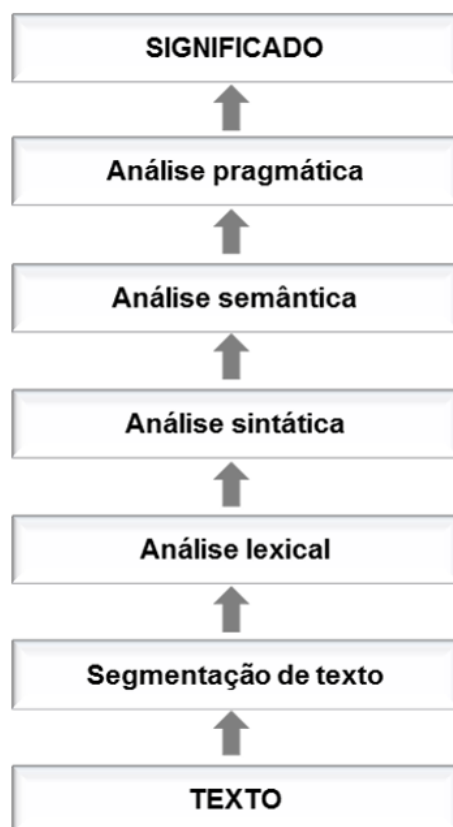


Figura 2.1: Etapas na análise de PLN (Adaptado) (Dale, 2010).

### 2.1.1 Segmentação de texto

A primeira tarefa a realizar quando se trabalha em PLN é a segmentação de texto. Esta pode ser realizada a vários níveis e o texto pode ser dividido em parágrafos, frases, palavras, sílabas ou morfemas. A complexidade desta tarefa varia consoante a língua em que é executada. Por exemplo, em línguas asiáticas, onde não há um delimitador explícito (como o espaço em branco nas línguas indo-europeias), a segmentação de texto é mais complicada. Mesmo quando o espaço em branco é normalmente o delimitador explícito, ocorrem situações de exceção (Tufis, 2009). Também Palmer (2010) refere que é necessário definir claramente caracteres, palavras e frases em qualquer documento e que a definição dessas unidades apresenta diferentes desafios, dependendo do idioma que é processado. Esta não é uma tarefa trivial, especialmente quando se considera a variedade de línguas humanas e sistemas de escrita. Assim, para este autor as línguas naturais são ambíguas e os sistemas de escrita muitas vezes amplificam essas ambiguidades, podendo também originar ambiguidades adicionais.

Quando se trata de páginas da *World Wide Web* (WEB), a insuficiente produção de metadados, quando se aplicam algoritmos no reconhecimento visual destas páginas, torna-se mais eficaz quando realizada parcialmente em segmentos de páginas que descrevem um único evento, em vez de toda a página (Fragkou, 2013).

### 2.1.2 Análise lexical

As palavras são os blocos de construção de textos em linguagem natural. A análise lexical incide sobre as técnicas e mecanismos de análise de texto ao nível da palavra (Hippisley, 2010). Para Hanks (2013) cada palavra, de

acordo com o contexto, tem vários significados possíveis. Assim, novas terminologias estão constantemente a ser criadas, nomeadamente as que se relacionam com nomes próprios e palavras compostas.

A análise lexical antecede normalmente o processamento sintático (Ranchhod, 2001), permitindo reduzir a complexidade da análise sintática ao classificar e simplificar algumas das operações realizadas sobre os dados (Platzer, 2013).

Ranchhod (2001) coloca em evidência que, para o processamento automático de texto, é necessário efetuar um tratamento das unidades básicas das línguas, palavras simples e palavras compostas. Este tratamento pode ser realizado pelos dicionários eletrónicos. Este autor menciona ainda que alguns dos problemas que se podem solucionar são: analisar contrações ou resolver ambiguidades. Já Liddy (2001) refere que a interpretação e representação do significado das palavras individuais varia de acordo com a teoria semântica utilizada no sistema de PLN e uma única unidade lexical pode ser decomposta nas suas propriedades mais básicas.

### 2.1.3 Análise sintática

Em PLN pressupõe-se naturalmente que a unidade básica, cujo significado pode ser analisado, é a frase. Deste modo uma frase exprime uma proposição, uma ideia ou um pensamento, sobre o mundo real ou imaginário. Extrair o significado de uma frase é uma questão-chave. Frases não são, no entanto, apenas a sequência linear de palavras, sendo necessária a sua análise gramatical para que estas possam ser interpretadas (Dale, 2010). Liddy (2001) considera que este nível centra-se na análise das palavras de uma frase para obter a sua estrutura gramatical. A mesma opinião têm Chopra, Prashar e Sain (2013) ao entenderem que as palavras são transformadas numa estrutura gramatical em que se podem visualizar os relacionamentos que existem entre si.

Ljunglöf e Wirén (2010) referem que nesta fase se analisam sequências de palavras (tipicamente uma frase) para determinar a sua descrição estrutural de acordo com a gramática formal, mas destacam que na maioria das circunstâncias, isso não é um objetivo em si, mas sim um passo intermédio que pode ter como objetivo, por exemplo, a atribuição de um significado à frase. Neste sentido, o resultado é tipicamente uma estrutura hierárquica, sintática, destinada a ser interpretada semanticamente.

### 2.1.4 Análise semântica

A sintaxe e a semântica de uma língua encontram-se intimamente relacionadas, sendo possível deduzir informações sobre a semântica unicamente a partir de estruturas sintáticas. Complementarmente, a informação de estruturas semânticas permite aperfeiçoar as análises sintáticas (Agustini, 2006).

Identificar a estrutura sintática subjacente de uma sequência de palavras é a etapa inicial na determinação do significado de uma frase. A análise sintática fornece um objeto estruturado que é mais suscetível de posterior manipulação e interpretação. A análise semântica permite atribuir sentido às frases analisadas gramaticalmente (Dale, 2010). A semântica determina os possíveis significados de uma frase. Este nível de processamento pode incluir desambiguação semântica de palavras com múltiplos sentidos.

Goddard e Schalley (2010) consideram que a análise semântica engloba não só a análise do significado das palavras, mas também das expressões fixas, frases inteiras e expressões no contexto. Na prática, isso significa traduzir expressões originais em algum tipo de metalinguagem semântica. Também Chopra, Prashar e Sain (2013) referem que ao resultado da análise sintática é atribuído significado consoante uma frase tenha, ou não, sentido.

Reshamwala, Mishra e Pawar (2013) alegam que nesta etapa, ao construir uma representação dos objetos e ações que constituem uma frase, incluindo os detalhes fornecidos por adjetivos, advérbios e proposições, é possível reunir informações vitais para uma posterior análise pragmática.

### 2.1.5 Análise pragmática

A determinação do significado de uma frase não é muitas vezes suficiente para compreender uma dada realidade, tornando-se assim necessário contextualizar esse significado. O pragmatismo é uma tradição filosófica que teve origem nos Estados Unidos da América, no século XIX, e que é atualmente impulsionada pelas discussões sobre a dinâmica evolutiva em que os objetos se revestem de significado através da sua utilização em contextos comportamentais (Artmann, 2013).

A análise pragmática relaciona a língua com a sua utilização. Deste modo, este nível utiliza o contexto, para além do conteúdo do texto, para entender o objetivo. Isto requer o conhecimento do mundo. Algumas aplicações do PLN podem, nestas circunstâncias, utilizar bases de conhecimento e módulos de inferência, de acordo com Liddy (2001). Esta autora refere ainda que podem ser utilizados diversos métodos para realizar a desambiguação, alguns que requerem informações sobre a frequência com que cada sentido ocorre num *corpus* de interesse particular, ou de uso geral, sendo que uns exigem a consideração do contexto local, e outros utilizam o conhecimento pragmático do domínio do documento.

Reshamwala et al. (2013, p. 114) definem pragmática como *"the analysis of the real meaning of an utterance in a human language, by disambiguating and contextualizing the utterance"* e destacam o papel importante que a identificação e eliminação de ambiguidades tem na interpretação do significado real das expressões utilizadas pela linguagem humana. Assim, quando uma palavra tem mais do que um significado, o seu sentido correto é encontrado através do conhecimento pragmático do contexto em que esta se insere.



# 3

## Extração automática

As preocupações com os aspetos pragmáticos da linguagem surgem da observação de que a comunicação em linguagem natural depende de circunstâncias que ultrapassam a linguagem propriamente dita. Deste modo, quando se participa num diálogo, ou se interpreta um texto, um vídeo ou uma imagem, é preciso fazer o seu enquadramento num determinado contexto. Assim, independentemente do tipo de informação, bem como do seu enquadramento, a extração automática de conteúdos tem um papel fundamental no PLN.

A extração automática pode ser realizada não só sobre informação textual, mas também sobre figuras, como um gráfico de linhas ou uma imagem de satélite, ou conteúdos multimédia, como um filme.

Choudhury e Giles (2015) constataam que muitas das descobertas experimentais, em documentos académicos, possuem figuras. Estas podem ser geradas a partir de dados que não constam desses documentos e conter informação só possível de identificar após a sua análise. A identificação do tipo de figura é o primeiro objetivo a ser realizado por estes autores. Assim, considerando que os gráficos de linhas são uma das figuras abundantes em trabalhos académicos, o foco principal do trabalho realizado por estes autores foi melhorar o módulo de processamento de imagem para realizar a extração de dados, totalmente automática, a partir de figuras que representem gráficos de linhas. Deste modo, a análise foi feita sobre documentos PDF e através de um algoritmo de segmentação de imagem foram identificadas regiões de texto e de imagem. Ao desenvolverem um

classificador binário de imagens, tornaram possível saber se uma dada imagem é um gráfico de linhas que será posteriormente analisado para extração dos seus dados.

A pesquisa semântica e extração de conteúdos multimédia revestem-se de interesse nas aplicações que apresentam vídeos. Chaudhari e Vankudothu (2015) desenvolveram uma metodologia que utiliza uma ontologia de domínio que aplica uma extração de conceitos que é depois utilizada para a extração de eventos. Outros autores (Manju & Valarmathie, 2015) referem que o aumento dos recursos multimédia, nomeadamente vídeos, leva a que se desenvolvam métodos inteligentes que permitam o seu processamento e organização. Sugerem assim que esse processamento comece com uma extração de objetos a partir dos conteúdos dos vídeos. Os objetos assim extraídos são também eles processados para extrair relações temporais. Este modelo, para além de extrair objetos e eventos, permite ainda associá-los automaticamente.

Bao, Tian, Chen e Lin (2015) evidenciam que a aquisição de dados florestais fornece um contributo importante para a construção de modelos ecológicos. Devido aos cálculos complexos e grandes volumes de dados associados com imagens de alta resolução espacial, a extração precisa e eficaz das copas das árvores individuais continua a ser um desafio. A abordagem proposta é validada comparando os seus resultados com os da aplicação de quatro algoritmos propostos por outros autores.

### 3.1 Extração automática de informação textual

Porfirio e Bidarra (2006) centram-se na complexidade da interpretação de textos, apontando alguns caminhos a seguir para um bom desenvolvimento deste tipo de interpretação e citando alguns métodos que são utilizados. Estes autores definem extração de informação como um método que consiste na identificação e manipulação de alguns aspetos linguísticos, tais como lexicais, sintáticos e semânticos, contidos nos textos. Por outro lado, Maybury (2008, p. 3) define extração de informação com mais pormenor ao escrever: *“Information extraction is the automated identification of specific semantic elements within a text such as the entities (e.g., people, organizations, locations), properties or attributes such as characteristics of entities, relations (among entities), and/or events”*. Este autor menciona que os sistemas que estudou são capazes de extrair entidades referidas em notícias com mais de 90% de precisão e as relações entre as entidades com uma precisão de 70-80%.

Em 2010, Savoy e Gaussier concluíram que os algoritmos de indexação automática estão muitas vezes associados aos mecanismos de extração de informação, nomeadamente a análise de estrutura e *tokenization*, a normalização morfológica, a atribuição de ponderações ou a remoção de *stopwords*. Neste último mecanismo, as palavras identificadas como *stopwords* não intervêm na interpretação do contexto, representando ruído no processo de recuperação podendo prejudicar o seu desempenho. Os autores destacam ainda que a remoção destas palavras permite reduzir o tamanho do armazenamento dos dados recuperados. Também a redução das palavras aos seus radicais ou raízes, técnica conhecida como *stemming*, pode aumentar a taxa de sucesso na extração automática de documentos. Nalgumas situações a sua aplicação pode ter de ser feita mais cuidadosamente, para evitar que o resultado não agregue no mesmo radical palavras com significados completamente diferentes (Savoy & Gaussier, 2010).

Gaizauskas, Saggion e Barker (2007) apontam diversos sistemas como tecnologias de acesso à informação, um dos quais o sistema de extração de informação. Uma das questões que estes autores pretendiam ver respondida no decurso do seu estudo era: *“How can human language technology assist a journalist to access the vast amount of information in a news archive? in particular can recent advances in NLP technologies, in areas such as question answering, summarisation, and information extraction, offer advantages in gathering background that standard information retrieval cannot?”* (Gaizauskas et al., 2007, p. 89). Para responder a esta questão, conceberam e implementaram um protótipo que utilizava tecnologia de extração de informações para apartar representações estruturadas de eventos, que por sua vez eram usados para preencher uma base de dados que posteriormente permitiria apoiar a pesquisa de eventos semelhantes. Estas tecnologias de acesso a infor-

mação eram incorporadas numa interface gráfica que possibilitava aos utilizadores procurar interativamente a informação que desejavam.

Para Wiebe e Riloff (2011) os sistemas de extração da informação envolvem tipicamente a identificação automática e a extração de informações factuais relativas a eventos. Estes sistemas podem assim extrair factos que se encontrem associados, desde incidentes terroristas a epidemias. Estes autores contrapõem os sistemas de extração de informação à análise de subjetividade, sendo a principal contribuição do seu trabalho a utilização de padrões de extração como uma representação complexa para expressões subjetivas. Consideram que a análise de subjetividade é uma área em rápido crescimento, que envolve a identificação automática e a extração de informações relativas a atitudes, opiniões e sentimentos a partir de textos não estruturados.

Liu (2010) refere que uma frase é objetiva se expressar informação factual sobre o mundo. Por outro lado, a subjetividade de uma frase é expressa através de sentimentos pessoais e convicções. Assim, as informações textuais podem ser classificadas em dois tipos principais: factos e opiniões. Factos são expressões objetivas sobre entidades, eventos e suas propriedades. Opiniões são geralmente expressões que descrevem sentimentos subjetivos das pessoas, avaliações ou sentimentos em relação a entidades, eventos e suas propriedades.

A utilização da análise de subjetividade permite melhorar a precisão dos factos, daí que seja importante o desenvolvimento de técnicas de extração que permitam aprender e reconhecer a linguagem subjetiva, que expressa opiniões e emoções, garantindo uma melhor performance dos sistemas de extração de informação (Wiebe & Riloff, 2011).

### 3.1.1 Áreas de aplicação

São muitas as áreas de investigação com trabalhos apresentados na extração automática de informação textual, desde a medicina, passando pelo ambiente até à história.

Kreuzthaler, Schulz e Berghold (2015) referem que da documentação da rotina diária dos ensaios clínicos faz parte grande volume de dados constituídos por textos livremente escritos em que a extração precisa e atempada de informação textual pode ser fulcral para a determinação de novos diagnósticos. À semelhança destes autores, também Mala e Lobiyal (2015) mencionam o grande volume de dados biomédicos que nesta área de investigação é necessário manusear. Estes autores referem que a maioria das informações são armazenadas em formato texto e as ontologias detêm um papel de destaque na extração de conceitos e sucessiva determinação de *clusters*. Outros autores (Moen et al., 2016) mencionam que uma das principais fontes de informações disponíveis em registos clínicos são notas de texto livre que documentam o atendimento ao paciente. Administrar essa informação pode ser bastante moroso e a sumarização automática de texto, criando resumos desses textos livres, pode ser um modo mais célere de obter respostas. Moen et al. (2016) avaliaram a eficácia de uma abordagem de avaliação automática experimental que incorporava medidas disponibilizadas pela aplicação ROUGE<sup>1</sup>. Este *software* desenvolvido para a avaliação automática de resumos mostrou-se viável face a uma avaliação humana de registos clínicos.

A construção de uma ontologia para auxiliar o processo de extração de informação espacial e temporal em notícias *online* relativas a ocorrências de riscos naturais é um trabalho que se encontra a ser desenvolvido por Wang e Stewart (2015). Deste modo, ao extrair informações semânticas de documentos, será possível determinar onde, quando ou como os eventos ocorreram. Assim, os utilizadores podem visualizar diretamente as tendências espaciais e temporais, analisar vários níveis de detalhe e observar aspetos que não são explicitamente descritos nos documentos de texto, permitindo relacionar múltiplas perspetivas de risco, nomeadamente ambientais e humanas. Os autores referem ainda que a metodologia utilizada neste trabalho para a ocorrência de riscos naturais pode ser aplicada em outro tipo de ocorrências, por exemplo, na extração semântica de eventos

---

<sup>1</sup><http://www.berouge.com>

de surtos de doenças ou campanhas políticas. A ontologia criada, utilizando a ferramenta *open-source* NeOn Toolkit<sup>2</sup>, inclui palavras-chave relativas à temática dos riscos naturais. Para além de comparar resultados, foi realizada uma avaliação para testar a associação automática de eventos extraídos relativamente à sua localização temporal e respetiva informação semântica. Fragkou (2013) relata a importância da informação semântica e considera que identificar segmentos de texto que contenham esta informação é um dos propósitos da extração de informação.

Horch, Kett e Weisbecker (2015) utilizam também uma aplicação *open-source*, Python<sup>3</sup>, para implementar o estudo que realizaram e avaliaram em 300 produtos, disponíveis para comprar *online*. A comparação de produtos, bem como dos seus preços, é uma tarefa essencial para quem compra e para quem vende. Neste trabalho os produtos são comparados com base no cálculo da similaridade entre pares dos seus atributos. Mas nem todas as consultas realizadas na WEB são acedidas diretamente. Muitas páginas estão escondidas e as suas informações tem de ser extraídas de forma dinâmica. Das e Kumar (2014) implementaram um sistema para que fosse possível preencher e enviar formulários da WEB automaticamente de acordo com uma determinada consulta do utilizador, confrontada com a ontologia de extração correspondente. Assim, o sistema extrai informação a partir das páginas, coloca os registos extraídos numa base de dados e interroga a base de dados que tem a consulta inicial para obter os dados relevantes por trás dos formulários da WEB.

Tendo como ponto de partida informação também disponível *online*, Wu, Liu e Fan (2015) constatarem que o conteúdo principal de uma página WEB é muitas vezes acompanhado por uma grande quantidade de conteúdo adicional. Para extrair conhecimento útil a partir do conteúdo principal, a informação adicional deve ser tratada como ruído e, como tal, removida. A resolução deste problema passa pela seleção de nós da árvore *Document Object Model* (DOM). Esta é uma interface, desenvolvida pelo *World Wide Web Consortium* (W3C) que permite aceder e atualizar o conteúdo, estrutura e estilo de documentos de forma dinâmica (W3C, 2005). Neste estudo considera-se numa primeira etapa a posição dos elementos na árvore, as suas áreas, fontes, textos, etiquetas e hiperligações para atribuir uma pontuação que representará a probabilidade de os nós pertencerem ao conteúdo principal. Numa segunda etapa agrupam-se os nós candidatos e removem-se os grupos que forem considerados ruído.

Atualmente são muitas as organizações que utilizam serviços de armazenamento em nuvem, mas ainda existem dificuldades em monitorizar as métricas de desempenho referenciadas nos contratos estabelecidos entre utilizadores e fornecedores deste tipo de serviço. Os contratos legais são normalmente grandes documentos de texto, com muitas expressões legais, exigindo muitas vezes a sua interpretação por parte de especialistas. Mittal, Joshi, Pearce e Joshi (2016) aplicam técnicas de mineração de texto para extrair automaticamente definições de termos chave e métricas que permitam monitorizar esses contratos.

Patil, Pharande, Nale e Agrawal (2015) destacam a existência de muitos documentos em formato digital que não têm resumos associados. Sumariar um texto automaticamente permite obter esses resumos. Para essa sumarização é necessária análise semântica, processamento de discurso, e interpretação inferencial através do agrupamento dos conteúdos. No seu trabalho, os autores acima referidos calculam a importância de cada frase no documento e extraem as frases mais importantes que irão originar o resumo, permitindo preservar o significado global do documento original.

Numa metodologia construída num sistema de pergunta-resposta para análise e processamento de documentos em português, Quaresma, Rodrigues, Prolo e Vieira (2006) apresentam um módulo de extração. Neste módulo referem diversas fases: análise sintática, análise semântica e interpretação semântica e pragmática. Na primeira processa-se o documento e cria-se um novo com a representação sintática de cada frase. Na segunda fase reescreve-se o documento anterior dando origem a uma nova coleção de documentos que possuem uma estrutura para representação do discurso. Na última fase processam-se os documentos resultantes da fase an-

---

<sup>2</sup><http://neon-toolkit.org>

<sup>3</sup><https://www.python.org>



terior, completando uma base de dados onde se guardam factos como termos de uma tabela. Estes autores referem que, na extração da informação, a informação semântica pode ser ambígua. Daí que possa haver mais do que uma interpretação para a mesma frase. Esta é uma limitação identificada no sistema acima referenciado. Também para Lahbib, Bounhas e Slimani (2015) é fundamental determinar as fontes de ambiguidade que influenciem diretamente a extração de termos. Em línguas como o árabe, a extração de um conjunto de termos que representem um domínio específico permitirá uma melhoria dos resultados obtidos, em consultas efetuadas sobre esse domínio.

Pintus, Yang e Rushmeier (2015) consideram que a conservação digital de documentos históricos é fundamental, uma vez que estes se vão deteriorando ao longo dos anos, e só assim se torna possível disponibilizá-los ao público. A quantidade bem como a importância das informações contidas nestes manuscritos motiva o desenvolvimento de ferramentas que os permita explorar, ler e apreciar. Neste sentido, apresentam dois métodos automáticos para: i) estimar a altura do texto; ii) extrair linhas de textos segmentadas. Também Efremova, García, Zhang e Calders (2015) desenvolvem trabalho sobre documentos históricos. Estes autores apresentam duas abordagens para a extração automática das relações familiares (por exemplo: marido, esposa, mãe, viúva) em documentos notariais que descrevem eventos, tais como vendas de propriedades, partilha de heranças ou inventários. Os dados pessoais presentes nesses documentos podem ser obtidos a partir de diferentes fontes, não têm normalmente uma estrutura uniforme, e muitas vezes são apresentados em formato de texto livre. A extração das relações familiares a partir desses dados permitirá conectar pessoas em diferentes documentos e fontes, possibilitando a descoberta de padrões sociais. Numa primeira abordagem são identificados os nomes de pessoas num dado documento, são gerados todos os possíveis pares de nomes candidatos, prevendo-se posteriormente se estes estão relacionados uns com os outros, utilizando técnicas de classificação. Na segunda abordagem, aplicam um modelo oculto de Markov para registar cada palavra de um documento com uma etiqueta indicando se: i) é um nome; ii) é um descritor específico de relacionamento; iii) não é nenhuma das opções anteriores. Posteriormente, os nomes que se encontram ligados por descritores de relacionamento são analisados.

## 3.2 Métricas de avaliação

Manning, Raghavan e Schütze (2008) descrevem métricas de avaliação que podem ser utilizadas na recuperação de informação. Destas destacam-se três medidas:

- *Precision* (P) – fração dos documentos recuperados que são relevantes;
- *Recall* (R) – fração dos documentos relevantes que são recuperados;
- *F measure* – combinação das medidas *precision* e *recall*.

Estas medidas podem ser calculadas pelas fórmulas que se apresentam na Tabela 3.1 (Manning et al., 2008).

Tabela 3.1: Formulação das métricas de avaliação.

<i>Precision</i>	$\frac{\#(\text{itens relevantes recuperados})}{\#(\text{itens recuperados})}$ $P(\text{relevantes} \text{recuperados})$
<i>Recall</i>	$\frac{\#(\text{itens relevantes recuperados})}{\#(\text{itens relevantes})}$ $P(\text{recuperados} \text{relevantes})$
<i>F measure</i>	$2PR/(P+R)$

A medida *F measure* permite medir o desempenho da extração efetuada. Tratando-se de uma média harmónica, entra em ponderação com o valor de *Recall* e *Precision*. Quanto mais próximo de 1, melhor o desempenho da extração (Manning et al., 2008).

Estas métricas são utilizadas em muitas das recentes investigações que envolvem extração automática de informação textual de modo a avaliar o trabalho desenvolvido. Apresenta-se na Tabela 3.2 algumas das publicações onde estas medidas foram aplicadas. Pela observação desta tabela, podemos constatar que, das 21 publicações, 12 (57%) referem a aplicação das três métricas referenciadas.

Tabela 3.2: Publicações vs métricas de avaliação.

Publicação	P	R	F
Acquisition of Generic Problem Solving Knowledge through Information Extraction and Pattern Mining (Halioui, Valtchev, & Diallo, 2015)	X	X	X
Arabic Terminology Extraction and Enrichment Based on Domain-Specific Text Mining (Lahbib et al., 2015)	X	X	X
ATHENA: Automatic Text Height Extraction for the Analysis of Text Lines in Old Handwritten Manuscripts (Pintus et al., 2015)	X	X	
Automatic extraction of correlation-entropy features for text document analysis directly in run-length compressed domain (Javed, Nagabhushan, & Chaudhuri, 2015)	X	X	
Automatic Web Content Extraction by Combination of Learning and Grouping (Wu et al., 2015)	X	X	X
Comparison of automatic summarisation methods for clinical free text notes (Moen et al., 2016)	X	X	X
Design of automatic extraction algorithm of knowledge points for MOOCs (Chen, Han, Dai, & Zhao, 2015)	X	X	X
Domain-specific keyphrase extraction and near-duplicate article detection based on ontology (Nhon Do & LongVan Ho, 2015)	X	X	
Extracting information from the text of electronic medical records to improve case detection: a systematic review (Ford, Carroll, Smith, Scott, & Cassell, 2016)	X	X	X
Extracting product unit attributes from product offers by using an ontology (Horch et al., 2015)	X	X	
Framework for automatic information extraction from research papers on nanocrystal devices (Dieb, Yoshioka, Hara, & Newton, 2015)	X	X	X
Identification of Query Forms for Retrieving the Information From Deep Web (Das & Kumar, 2014)	X	X	
Large-scale automatic extraction of side effects associated with targeted anticancer drugs from full-text oncological articles (Xu & Wang, 2015)	X	X	X
litewi: A combined term extraction and entity linking method for eliciting educational ontologies from textbooks (Conde, Larrañaga, Arruarte, Elorriaga, & Roth, 2016)	X	X	X
Morphology based text detection and extraction from Malayalam news vídeos (Anoop & Lajish, 2015)	X	X	
Optimizing Ontology Alignments through NSGA-II Using an Aggregation Strategy and a Mapping Extraction Approach (Xue & Jiang, 2015)	X	X	X
Organizing multimedia big data using semantic based video content extraction technique (Manju & Valarmathie, 2015)	X	X	
Secondary use of electronic health records for building cohort studies through top-down information extraction (Kreuzthaler et al., 2015)	X	X	X
Spatiotemporal and semantic information extraction from Web news reports about natural hazards (Wang & Stewart, 2015)	X	X	
Stacked Ensembles of Information Extractors for Knowledge-Base Population (Viswanathan, Rajani, Bentor, & Mooney, 2015)	X	X	X
Supporting the semi-automatic semantic annotation of web services: A systematic literature review (Tosi & Morasca, 2015)	X	X	

# 4

## Ontologias

A utilização de ontologias conjuntamente com o PLN é uma ferramenta importante na construção de conhecimento, permitindo obter novas respostas nas áreas de estudo mais diversas, sendo a história apenas um exemplo daquelas.

### 4.1 Definir

Um dos desafios das tecnologias de informação é fazer chegar aos utilizadores a informação certa, no momento certo (Uschold & Gruninger, 2004). Este é um desafio que se deseja atingir ao construir uma ontologia. Com a disponibilização em tempo real de qualquer tipo de informação em numerosos repositórios de dados, a dificuldade não é a ausência de informação, mas sim como extrair somente o que é mais relevante face ao objetivo em estudo (Aundhakar & Pokale, 2015). A criação de ontologias pode ser um caminho para melhor organizar-se a informação e, deste modo, extrair aquela que é mais importante para um determinado estudo. Esta organização da informação, conseguida através de uma ontologia, poderá ser valorizada, gerando conhecimento que pode ser posteriormente partilhado.

Ao reconhecer a importância das ontologias em diversas áreas de aplicação, Guarino (1998) destacou dois tipos de peculiaridades. O primeiro considera as arquiteturas escolhidas, que evidenciam o papel central que

as ontologias podem tomar num determinado sistema de informação. O segundo diz respeito às metodologias aplicadas, salientando-se abordagens interdisciplinares. Assim, desde o início dos anos 90 que o conceito de ontologia ganha popularidade e aparece associado a vários domínios de aplicação, nomeadamente: engenharia do conhecimento, PLN e representação do conhecimento, construção de bases de dados, educação, extração e recuperação de informação (Baoxian & Yang, 2013; Bajenaru & Smeureanu, 2015).

Tawfik, Giunchiglia e Maltese (2014) referem que um domínio específico de interesse pode ser compreendido através de ontologias. Posteriormente pode ser partilhado entre utilizadores e ser a base para tornar o raciocínio automático em diversas aplicações.

Inicialmente, a palavra "ontologia" aparecia apenas referida em contexto filosófico, referindo-se à natureza do ser ou ao seu tipo de existência (Van Heijst et al., 1997). Pretorius (2004), embora considere ontologias no sentido filosófico, acrescenta que estas são sistemas de categorias, quadros de referência, que representam uma certa visão do mundo. Gruber (2009) concorda com a natureza filosófica deste conceito, acrescentando que ontologia é um termo técnico que designa um artefacto projetado para uma dada finalidade: permitir a modelagem de conhecimento sobre um domínio, real ou imaginário. Para Guarino, Oberle e Staab (2009) a palavra "ontologia" é aplicada em diferentes contextos, com diferentes sentidos, por diferentes utilizadores. Também estes autores escrevem sobre o sentido filosófico que lhe é atribuído, o qual apresenta uma tradição bem estabelecida, em oposição ao sentido computacional, bem mais recente. Assim, num contexto computacional, as ontologias são um meio para modelar formalmente a estrutura de um sistema, para que se consiga atingir um determinado propósito.

Ao longo dos anos, são diversos os autores que referem a importância da partilha de informação num determinado domínio, desde Noy e McGuinness (2001), que afirmam que uma ontologia desenvolve um vocabulário comum para utilizadores que partilham informação num dado domínio, até Mishra e Jain (2015), que apoiam esta opinião ao referirem que as ontologias representam e partilham conhecimento num domínio de aplicação.

Para Shah, Kongari e Dedhia (2015) uma ontologia permite definir um domínio com enfoque na partilha de informação, compartimentando as variáveis necessárias para um conjunto de cálculos, estabelecendo relações entre elas. Também Mukhopadhyay e Shikalgar (2013) partilham desta opinião ao referirem que uma ontologia descreve os conceitos do domínio e as relações que são mantidas entre esses conceitos. Uma ontologia, é, assim, uma base de conhecimento disponibilizada para diferentes aplicações que necessitem de a utilizar e/ou partilhar (Bajenaru & Smeureanu, 2015). Outros autores vão mais longe ao mencionarem o interesse de existir um compromisso de partilha no que diz respeito à utilização de ontologias. Sem este compromisso, os benefícios da utilização de ontologia são limitados (Guarino et al., 2009).

Zhong, Ding, Love e Luo (2015) consideram que uma ontologia é a representação formal do abstrato. Complementam, no entanto, esta afirmação escrevendo: *"Ontologies are flexible data structures that can be changed and adapted to enable the semantics of planned knowledge to be interpreted in a formal way using a computer"* (Zhong et al., 2015, p. 48). No contexto das bases de dados, também Gruber (2009) refere que uma ontologia pode ser vista como um nível de abstração, mas relativamente ao modelo de dados.

Foram muitas as definições de "ontologia" encontradas na bibliografia consultada (e.g. (Guarino, 1998); (Noy & McGuinness, 2001); (Uschold & Gruninger, 2004); (Antoniou & Van Harmelen, 2008)). No entanto, muitos são os autores que referem a definição proposta por Thomas Gruber (1993, p. 1): *"An ontology is an explicit specification of a conceptualization"*. Para este autor, o que existe é o que pode ser representado, devendo uma ontologia permitir a definição de classes, propriedades e relacionamentos entre classes.

Uschold e Gruninger (2004) concretizam algumas das referências anteriormente apresentadas ao mencionarem que as ontologias representam muitos tipos diferentes de "coisas", numa determinada área. Estas "coisas" são representadas em classes e subclasses. Cada classe é tipicamente associada a várias propriedades que descrevem as suas características, atributos e restrições. Para esses autores, uma ontologia constitui uma

base de conhecimento.

Em 1995, Gruber destaca a clareza, objetividade, coerência e extensibilidade que o desenvolvimento de uma ontologia deve apresentar. Assim, para este autor, ao construir uma ontologia dever-se-á garantir:

- definições claras e objetivas, documentadas em linguagem natural e independentes do contexto social ou computacional;
- uma estrutura coerente, sem contradições, e logicamente consistente;
- uma estrutura extensível, que preveja a utilização de vocabulário partilhado, antecipando questões futuras;
- que a adição de termos à estrutura da ontologia não levará a uma modificação do conteúdo já existente;
- o menor número possível de referências à realidade a modelar, de modo a garantir a especialização noutras áreas.

Noy e McGuinness (2001) consideram que não há uma metodologia "correta" para o desenvolvimento de ontologias. Estes autores apontam, ainda assim, três regras que consideram fundamentais: não há uma maneira única de modelar um domínio – existem sempre outras alternativas; o desenvolvimento de ontologias é sempre um processo iterativo; os conceitos na ontologia devem ser semelhantes a objetos e relações no domínio de interesse. São diretivas muito abertas que se irão seguir.

## 4.2 Classificar

Ao construir uma ontologia esta é enquadrada num determinado tipo. Apresenta-se na Tabela 4.1 uma classificação possível dos diversos tipos de ontologias.

Tabela 4.1: Tipos de ontologias (Adaptado) (Guarino, 1998; Guarino et al., 2009).

Tipo de ontologia	Descrição
Ontologias de nível superior	Descrevem conceitos muito gerais (como por exemplo: espaço, tempo, objeto, evento) que são independentes de um problema ou domínio em particular.
Ontologias de domínio	Representam o vocabulário relevante num determinado tipo de domínio, por exemplo, histórico.
Ontologias de tarefa	Descrevem uma tarefa genérica ou atividade, por exemplo, vendas.
Ontologias de aplicação	Reproduzem conceitos dependentes de um domínio e de uma tarefa, sendo muitas vezes especializações de ontologias relacionadas.

No estudo apresentado nesta tese as ontologias construídas representam conhecimento sobre dados históricos de natureza prosopográfica, correspondendo a uma ontologia de domínio. É, portanto, especializada.

Também Ide e Woolner (2007) apresentam uma classificação para ontologias. Assim, uma ontologia histórica pode ser retratada segundo duas perspetivas: sincrónica e diacrónica. A primeira representa o estado do mundo durante um determinado intervalo de tempo e a segunda representa as mudanças que se vão desenrolando ao longo do tempo. A perspetiva sincrónica é fornecida por uma série de ontologias instantâneas que são linearmente ordenadas ao longo de um período de tempo. A perspetiva diacrónica reúne a ontologia de eventos que abrange todo o período de tempo coberto pela série de ontologias instantâneas.

Neste sentido, e relativamente aos dados utilizados para dar resposta às nossas questões, pode-se, por um lado, considerar uma perspetiva diacrónica quando se realizam referências a toda a época em que decorreu a

observação em torno do SO (1570-1773); por outro lado, pode-se atender a uma perspectiva sincrónica quando se faz referência apenas a um intervalo dessa época.

Valaski, Malucelli e Reinehr (2012) mencionam que, independentemente do tipo de ontologia, a sua aplicação tornou-se uma referência em diversas áreas e aplicações, uma vez que permite representar o conhecimento através de estruturas simples ou complexas, com conceitos mais genéricos ou especializados. Assim, as ontologias são consideradas um importante recurso nas organizações, permitindo uma melhoria contínua na utilização, representação e interpretação do conhecimento. Elsayed, El-Beltagy, Rafea, e Hegazy (2007) reforçam esta ideia ao concluírem que o desenvolvimento de ontologias melhora a organização, gestão e compreensão das informações disponíveis, sendo fulcrais na partilha e reutilização de conhecimentos. A capacidade de lidar com grandes quantidades de informação torna bastante interessante o desenvolvimento de ontologias. Estas duas realidades que se acabam de referir são hoje fulcrais no conhecimento científico.

As ontologias são habitualmente classificadas de acordo com as suas características. Uma das classificações mais divulgadas é a desenvolvida por Guarino (1998). Segundo este autor, as ontologias são classificadas de acordo com o nível de generalização do conteúdo que representam: ontologias de alto nível; ontologias de domínio; ontologias de tarefa; ontologias de aplicação. Posteriormente, Gašević, Djuric e Devedžić (2006) classificaram as ontologias como formais e informais. Estes autores consideram que, independentemente da definição utilizada, as ontologias são parte fundamental do conhecimento. Já Antoniou e Van Harmelen (2008) estreitam a definição de ontologia, considerando que uma ontologia descreve formalmente um domínio do discurso e é composta por uma lista finita de termos e das respetivas relações entre esses termos. Os termos referem-se a conceitos importantes (classes de objetos) do domínio. Por exemplo, num ambiente académico, funcionários, alunos, cursos e disciplinas são alguns conceitos importantes.

No âmbito do nosso estudo, ocupações (tal como se definiam na época), estatutos e setores de atividades, são alguns exemplos de conceitos considerados fundamentais para a classificação dos dados. Estabelecer estes conceitos, muitas vezes com recurso a arquivos digitais ou analógicos, bem como consultas a especialistas, permitiu criar uma grelha para utilizar na classificação dos eventos.

### 4.3 Construir

*"Why develop an ontology?"* Esta é a questão com que Noy e McGuinness (2001, p. 1) iniciam o seu artigo. Ao tentar responder a esta pergunta, apontam como razões para a construção de uma ontologia:

- a partilha, entre os utilizadores, do conhecimento relativo a um dado domínio;
- a reutilização de conhecimento numa dada área de interesse;
- a criação de pressupostos de domínio explícitos que possam ser facilmente alteráveis;
- a separação do conhecimento do domínio do conhecimento operacional;
- a análise do conhecimento de domínio.

O desenvolvimento das duas ontologias construídas foi executado recorrendo à metodologia apresentada por Noy e McGuinness (2001). Dhingra e Bhatia (2015) e Fraihat e Shambour (2015) são alguns dos autores que referem a abordagem dos sete passos proposta por Noy e McGuinness (2001):

1. Determinar o domínio e o âmbito da ontologia: que domínio irá a ontologia conter? No que é que vai ser utilizada a ontologia? A que tipo de perguntas deve a ontologia dar resposta? Quem vai utilizar e como é feita a gestão da ontologia?

2. Ponderar a reutilização de ontologias;
3. Enumerar os termos mais importantes da ontologia: criar uma lista detalhada de termos onde não é necessário verificar se existe sobreposição entre conceitos, relações e/ou propriedades;
4. Definir classes e a sua hierarquia: duas metodologias diferentes (top-down: criar as classes mais gerais; bottom-up: definir as classes mais específicas);
5. Definir propriedades das classes: diversos tipos (intrínsecas, extrínsecas, de relação...);
6. Definir as características das propriedades: cardinalidade, tipo de valor ...;
7. Criar instâncias individuais das classes.

Considerou-se, para a construção da ontologia, algumas destas sugestões, nomeadamente quando começámos por escolher o seu domínio e aplicámos na definição da hierarquia de classes uma metodologia *top-down*.

## 4.4 Linguagens

A construção de uma ontologia está sempre ligada ao conceito de linguagem. Assim, são as linguagens que permitem representar as ontologias.

Decker et al. (2000) previam que na próxima geração da WEB as informações disponíveis deixariam de ser apenas destinadas à interpretação humana e teriam um papel fundamental no processamento computacional. Esta previsão, que se viria a confirmar, referia ainda que a WEB semântica requeria interoperabilidade ao nível sintático e semântico dos documentos analisados. Assim, o termo WEB semântica está associado à interligação de dados e de utilizadores na internet, permitindo que, através de computadores, seja possível desenvolver sistemas que permitam não só criar armazenamento de dados mas também construir vocabulário e escrever regras, que ao manipular os dados permitam a aquisição de novos conhecimentos (W3C, nd).

Linguagens como *Resource Description Framework* (RDF) e *Web Ontology Language* (OWL) contribuem para o desenho das ontologias, permitindo interligar os dados que estas representam. Associada a estas duas linguagens encontra-se a *Simple Protocol and RDF Query Language* (SPARQL), considerada a linguagem de consulta *standard* pela comunidade da WEB semântica (Calvanese et al., 2016). Para Pérez, Arenas e Gutierrez (2006) a formalização desta linguagem traz diversos benefícios, nomeadamente a possibilidade de encontrar novas relações e identificar redundâncias e contradições.

Antoniou e Van Harmelen (2004) afirmam que a linguagem RDF é deliberadamente limitada a uma hierarquia de classes e propriedades com definição de domínios e contradomínios. Tratando-se de um modelo padronizado para o intercâmbio de dados, esta linguagem tem características que facilitam a interligação e partilha de dados entre diversas aplicações, sendo a sua estrutura fundamental a *triple* constituída por sujeito, predicado e objeto (W3C, 2014).

Baseada em lógica computacional, a OWL é uma linguagem concebida para representar conhecimento sobre "coisas", grupos de "coisas" e relações entre "coisas". A versão atual desta linguagem surgiu em 2009, com uma segunda edição em 2012, tendo por base uma primeira versão datada de 2004 (W3C, 2013). Qualquer uma delas foi desenvolvida por grupos de trabalho do consórcio W3C. A OWL é atualmente uma linguagem *standard* no desenho e construção de ontologias. Baseia-se nos conceitos sintáticos da linguagem RDF, utilizando as suas primitivas de modelação (Antoniou & Van Harmelen, 2004).

Kollia, Glimm e Horrocks (2011) referem que a possibilidade de "interrogar" é fundamental no contexto da WEB semântica. Deste modo, a existência de mecanismos através dos quais os utilizadores e as aplicações possam

interagir com ontologias e dados são essenciais.

Com o aparecimento da linguagem RDF foi necessário encontrar novas abordagens que permitissem uma diminuição do tempo de execução das consultas realizadas sobre os dados (Hartig & Heese, 2007). A SPARQL é a linguagem utilizada em muitas das consultas realizadas no âmbito da WEB semântica.

Na sua terminologia, a SPARQL inclui diversos termos dos quais se destacam os descritos na Tabela 4.2.

Tabela 4.2: Terminologia da SPARQL (Adaptado) (W3C, 2008).

Termo	Descrição
<i>IRI</i>	<i>String unicode</i> sem quaisquer caracteres de controlo.
<i>Literal</i>	Identificação de valores tais como datas e números, por meio de uma representação lexical.
<i>Lexical form</i>	<i>String unicode</i> normalizada.
<i>Plain literal</i>	Combinação de uma <i>string</i> com uma tag de idioma opcional.
<i>Typed literal</i>	Combinação de uma <i>string</i> com tipo de dados.
<i>Datatype IRI</i>	Tipo de dados de uma <i>string unicode</i> .

Hazber, Li, Gu, Xu, e Li (2015) consideram que o desenvolvimento da WEB semântica pode passar pela construção de ontologias, e respetivas consultas, a partir de bases de dados relacionais. Na metodologia que propõem reescrevem as consultas criadas em SLQ em consultas SPARQL. Reescrever consultas foi igualmente o objetivo do trabalho apresentado por Hartig e Heese (2007), que através de algoritmos transformaram consultas SPARQL em *SPARQL query graph model* (SQGM). Estes autores pretendiam simplificar a escrita destas consultas, tornando-as mais eficientes. Pérez et al. (2006) referem também a vertente gráfica das consultas SPARQL, mencionando que esta linguagem permite realizar consultas através da combinação das representações gráficas dos dados. Estes autores dividiram as consultas em três partes. Na primeira parte, consideraram a possibilidade de aplicar filtros e escolher a fonte de dados utilizada, na segunda a aplicação de operadores e na última a escolha do tipo de resultado que a consulta produzirá.



# 5

## Recursos utilizados e metodologias aplicadas

### 5.1 Ferramentas de desenvolvimento

As tabelas necessárias para a classificação dos dados prosopográficos foram exportadas para o PostgreSQL 9.4<sup>1</sup>. Foi nesta aplicação que se fez a remoção de *stopwords* e aplicação de *stemming*. O PostgreSQL é um sistema de gestão de bases de dados relacionais (SGBDR) disponível gratuitamente para *download*. As suas origens remontam ao ano de 1986 com o projeto POSTGRES desenvolvido por uma equipa da Universidade da Califórnia. Este projeto passou por várias versões e em 1995 originou o PostgreSQL (PostgreSQL Global Development Group, 2014).

O consórcio W3C considera que não existe um editor "melhor" do que os outros e disponibiliza, em [http://www.w3.org/wiki/Ontology\\_editors](http://www.w3.org/wiki/Ontology_editors), uma lista de editores de ontologias, na sua maioria *open-source*.

Apresenta-se na Tabela 5.1 o endereço dos editores que, estando disponíveis para *download* gratuito (em 17 de setembro de 2015), podem ser consultados.

---

<sup>1</sup><https://www.postgresql.org>

Tabela 5.1: Editores de ontologias.

Designação	Produtor	Endereço	Versão
Protégé	Stanford University, Stanford Center for Biomedical Informatics Research	<a href="http://protege.stanford.edu">http://protege.stanford.edu</a>	5.0.0
NeOn Toolkit	The NeOn foundation	<a href="http://neon-toolkit.org">http://neon-toolkit.org</a>	2.5.2
Vitro	Cornell University	<a href="http://vitro.mannlib.cornell.edu">http://vitro.mannlib.cornell.edu</a>	1.8
OWLGrEd	University of Latvia, The Institute of Mathematics and Computer Science	<a href="http://owlgred.lumii.lv">http://owlgred.lumii.lv</a>	1.6.1
Semantic Turkey	Food and Agriculture Organization of the United Nations (FAO)/Tor Vergata, University of Rome	<a href="http://semanticturkey.uniroma2.it">http://semanticturkey.uniroma2.it</a>	0.12

Para a construção das ontologias foi necessário escolher um editor que permitisse a definição de classes, a implementação de propriedades, bem como o posterior povoamento das ontologias.

Dos editores referenciados na Tabela 5.1 não foram opção de escolha os editores NeOn Toolkit e Vitro, por não apresentarem versões tão atualizadas como as restantes ferramentas, tanto ao nível da aplicação como dos *plug-ins*.

Relativamente à aplicação OWLGrEd, tratando-se de um editor gráfico para ontologias, não contempla nas opções que disponibiliza a possibilidade de construção automática de uma ontologia a partir das tabelas de um SGBDR. Tem, no entanto, características adicionais, podendo ser um dos *plug-ins* possível de utilizar no editor Protégé.

Assim, dos editores mencionados na Tabela 5.1 apenas Semantic Turkey e Protégé continuam a ter assiduamente atualizações, tendo as últimas versões de ambas sido disponibilizadas em 2016.

Semantic Turkey é uma plataforma para a aquisição e gestão do conhecimento, desenvolvida pela universidade de Roma. Embora possua ferramentas que permitem desenvolver ontologias, não se encontra tão desenvolvida ao nível da conexão a SGBDR, nem disponibiliza uma diversidade tão elevada de *plug-ins*.

Deste modo, pela sua interoperabilidade com outras ferramentas, pelas versões em constante atualização e pela disponibilidade elevada de *plug-ins*, o editor Protégé foi a opção escolhida para o desenvolvimento das ontologias que se apresentam nos cenários I e II referidos na secção 6.

### 5.1.1 Protégé

O *software* Protégé 5.0 (Stanford Center for Biomedical Informatics Research, 2014) é disponibilizado gratuitamente<sup>2</sup>. Este editor, desenvolvido e gerido pela Universidade de Stanford, é sustentado pelos mais diversos utilizadores (como por exemplo: comunidades académicas e governamentais), que utilizam esta ferramenta para construir soluções baseadas no conhecimento em áreas tão diversas como a biomedicina ou a modelação organizacional. Através de diversas ferramentas, este *software* permite, por um lado, interpretar e processar dados, e por outro comparar e juntar ontologias (Kumar et al., 2013).

O Protégé possibilita a descrição de termos através da construção de taxonomias (Han et al., 2013; Ramkumar & Poorna, 2015). O desenvolvimento da ontologia assenta essencialmente na criação de classes, atribuição de propriedades a essas classes e aplicação de restrições às propriedades. Posteriormente, é possível gerar automaticamente interfaces. A criação de novas instâncias é realizada através dessas interfaces. Assim, para

<sup>2</sup><http://protege.stanford.edu>

cada classe é criado um formulário personalizável (Knublauch et al., 2004).

À semelhança de muitas outras ferramentas de modelação, a arquitetura do Protégé é separada em *model* e *view*. O primeiro é um mecanismo de representação interna para ontologias e bases de conhecimento. A segunda permite ao utilizador utilizar uma interface para visualizar e manipular o modelo subjacente (Knublauch et al., 2004).

Neste trabalho considerou-se a nomenclatura sugerida pelo *software* Protégé (Stanford Center for Biomedical Informatics Research, 2014). Assim sendo, em vez de propriedades de classe, (Noy & McGuinness, 2001) designam-se os relacionamentos entre classes como propriedades de objeto.

Apresenta-se no Apêndice A uma listagem de publicações, editadas entre janeiro de 2013 e julho de 2015, cuja temática aborda ontologias. Realizou-se uma análise a essas publicações para determinar quais as que referiam a ferramenta Protégé (independentemente da sua versão). Atribuiu-se o valor 0 às que não referem a ferramenta, o valor 1 para as que apenas a referenciam e o valor 2 para as que a utilizam nos seus ensaios. Assim, das 241 publicações consultadas, foi atribuído a 54 (22,4%) o valor 0, a 27 (11,2%) o valor 1 e a 160 (66,4%) o valor 2. Atribuiu-se também a cada publicação uma área de atuação. Nalgumas situações identificou-se mais do que uma área, mas optou-se por alocar cada uma das publicações somente à área que se considerou principal. Tratando-se de uma ferramenta desenvolvida inicialmente para a área da medicina, encontram-se atualmente estudos noutros campos. Salientamos, assim, o facto de ter-se encontrado estudos ontológicos em áreas tão diversas como o turismo, a arqueologia ou a energia.

## 5.2 Cenário de referência

Para a construção das ontologias utilizou-se como base de conhecimento os dados contidos no SPARES. Neste sistema, os dados históricos encontram-se distribuídos por diversas tabelas (Tabela 5.2). No entanto, pode-se considerar a tabela Comissários como central, uma vez que é nesta que se encontram as relações entre os diversos intervenientes do SO. Esta tabela é um exemplo da vantagem de uma "tabela poder ter mais do que uma finalidade e dos seus dados poderem ser vistos com diferentes formas e formatos" (Caldeira, 2011, p. 25).

Tabela 5.2: Descrição das tabelas principais da base de dados SPARES.

Tabelas do SPARES	Descrição
Comissários, aliás Eventos & relações	Dicionário dos eventos biográficos e relacionais dos indivíduos.
Ministros e oficiais	Identificação e caracterização dos ministros e oficiais do SO.
Actor	Identificação e caracterização dos indivíduos.
Ego	Identificação e caracterização dos indivíduos para efeitos de genealogia.
Genealogia	Identificação e caracterização das relações genealógicas entre os egos.
Localidade	Identificação e caracterização de localidades.
Documento	Contextualização dos eventos.
Arquivo	Contextualização dos documentos consultados.

Para além das tabelas referidas, o SPARES contém ainda tabelas de controlo da base de dados e tabelas auxiliares, necessárias para uma utilização mais eficiente das interfaces utilizadas para exploração dos dados.

Embora existam diversas tabelas na base de dados, o ponto de partida foi a tabela Comissários, com 161145 registos, em 19 de setembro de 2015. É nela que se inserem as relações entre os diversos indivíduos. Muitos desses indivíduos eram ministros e agentes locais da Inquisição (inquisidores, comissários, familiares, notários), ou Santo Ofício, ou pessoas que estavam com eles relacionadas (por exemplo: testemunha, réu). Tratava-se assim de um dicionário dos eventos (biográficos e relacionais) que envolvem esses indivíduos. Eventualmente

também surgem eventos de parentesco, que são sempre relacionais.

As personagens são designadas por P1 e P2. A leitura é sempre feita de P2 para P1, ou seja, cada pessoa P2 tem uma relação do tipo "Testemunha na habilitação SO" com um indivíduo P1. Neste caso, associado a esta relação está um evento de "Depoimento", do tipo "Relacional". Olival, Garcia, Lopes e Sequeira (2013) referem que este tipo de evento era na época quase uma obrigação. Uma notificação do SO para depor como testemunha levava muitas vezes a que, em vez de "fugir" a este tipo de depoimento, se cometesse perjúrio nos juramentos, mas a pessoa comparecia. Pode-se observar na tabela 5.3 o relacionamento atrás referido.

Tabela 5.3: Exemplo parcial de um registo retirado do SPARES.

Nome de P1	Relação	Nome de P2	Tipo de evento	Evento
*****	Testemunha na habilitação SO	*****	Relacional	<Depoimento>

Na base de dados podem observar-se 32 eventos. Estes podem ser consultados na Tabela 5.4, que apresenta uma breve descrição de cada um deles.

Tabela 5.4: Descrição de eventos.

Designação dos Eventos	Descrição
Agente institucional	Comissário, notário, deputado ou outro oficial ou ministro que atua num processo.
Autoria	Trabalho publicado pelo indivíduo.
Avaliação	Interlocutórias, votos a aprovar ou reprovar alguém.
Batismo	Data em que o indivíduo foi batizado.
Casamento	Regista quando uma pessoa se casa.
Comercial	Compra ou venda de bens e produtos, móveis ou imóveis.
Conhecimento	Etiqueta para identificar o evento de conhecimento entre duas pessoas.
Correspondência	Correspondência trocada.
Depoimento	Regista o que afirma judicial ou extra-judicialmente uma testemunha.
Estatuto social	Atributo de distinção de alguém ou posição na hierarquia social.
Falecimento	Etiqueta para identificar o evento de falecimento de uma pessoa.
Físico	Para descrições de atributos físicos.
Formação Académica	Regista o grau de formação académica.
Genealogia	Ascendentes conhecidos.
Herança	Testamenteiros ou outros dados sobre heranças.
Informação final do comissário	Relatório final do comissário numa habilitação.
Intermediário	Pessoa que medeia qualquer "negócio" ou diligência. Ex.: procurador, fiador, broker.
Local preciso	Local preciso da habilitação. Ex: igreja, capela, casas de residência, etc.
Mercê	Dádiva graciosa ou remuneratória feita a alguém.
Morador	Etiqueta para identificar o evento de morada, ou seja, uma pessoa habita em: 'Local'.
Nascimento	Etiqueta para identificar o evento de nascimento de uma pessoa.
Ocupação	Regista a ocupação de uma pessoa numa determinada altura da sua vida.
Pagamento	Entrega de valores a alguém a troco de alguma coisa.
Parceria	Associação em negócio comercial.
Parentesco	Grau de parentesco entre duas pessoas.
Processo	Diligência processual em que alguém está ou esteve envolvido.
Receita	Entrada de qualquer tipo de valor efetivo, mesmo que seja em géneros.
Relacionamento	Relações de amizade ou inimizade.
Rendimento ou capital	Valor da renda ou do capital que alguém dispõe.
Saber	Conhecimentos de alguém. Ex: sabe línguas.
Saúde	Estado de saúde.
Transporte	Relações emergentes de negócios que envolvem transporte de pessoas ou bens como fretar, carregar, financiar.

Apresenta-se na Tabela 5.5 os eventos possíveis de ocorrer entre P2 e P1, distribuídos por tipo de evento.

A construção das ontologias apresentadas neste trabalho incidirá sobre o evento ocupação, ou seja, o cargo ou profissão que alguém tinha. Constata-se, assim, por leitura da Tabela 5.5, o leque possível de eventos (biográficos, relacionais e de parentesco) que pautavam a vida das entidades estudadas.

Tabela 5.5: Designação dos eventos por tipo.

Designação dos eventos	Tipo de evento		
	Biográfico	Parentesco	Relacional
Agente institucional	X		X
Autoria	X		
Avaliação	X		X
Batismo	X	X	
Casamento	X		X
Comercial	X		X
Conhecimento		X	
Correspondência	X	X	X
Depoimento	X		X
Estatuto social	X		
Falecimento	X		
Físico (descrição do)	X		
Formação Acadêmica	X		
Genealogia	X		X
Herança	X		X
Informação final do comissário			X
Intermediário	X	X	X
Local preciso			X
Mercê	X		
Morador	X		X
Nascimento	X		
Ocupação	X		X
Pagamento	X		X
Parceria			X
Parentesco	X	X	X
Processo	X		X
Receita	X		X
Relacionamento	X		X
Rendimento ou capital	X		
Saber	X		
Saúde	X		
Transporte	X		X

Foi criada uma lista das designações da época para as diversas ocupações, de modo a agrupá-las. A partir destas elaborou-se uma listagem dos diversos setores de atividades (cruzando grandes áreas da época com classificações atuais). Esta classificação teve como ponto de partida um repositório de dados com cerca de 30000 registros, referentes a diversos indivíduos. Foi atribuído a cada um dos indivíduos um setor de ocupação, um setor de atividade e um estatuto. Foi a partir deste cenário de referência que se associou uma ocupação, atividade e estatuto aos eventos classificados.

Pela análise dos dados, constatou-se que as grandes áreas de ocupação da época podem corresponder a mais do que um setor de atividade, tal como foi acima definido. Na Figura 5.1 pode-se observar essa relação.

		Setor de atividade																
		Administração	Agricultura	Belas Artes	Defesa	Diplomacia	Doméstico	Eclesiástico	Ensino	Justiça	Mineração	Não se sabe	Navegação	Negócio	Pecuária	Pesca	Saúde	Segurança
Ocupação (designação da época)	Artes liberais	●							●	●							●	
	Artes mecânicas								●								●	
	Belas Artes			●														
	Embaixadas					●												
	Estudo								●									
	Governo da casa	●					●					●						
	Governo da res publica	●			●			●		●							●	●
	Igreja							●										
	Lavoura		●												●			
	Mercancia													●				
	Milícia				●												●	
	Mineração										●							
	Não se sabe	●										●						
	Navegação												●					
	Pesca															●		
	Sem ocupação																	●
	Sem ocupação/nobre																	●
	Viver da sua fazenda																	●

Figura 5.1: Agrupamento de ocupações (designação da época) por setores de atividade.

As ocupações encontram-se associadas a diferentes estatutos sociais. Apresenta-se na Tabela 5.6 os tipos de estatuto possíveis para cada ocupação/setor de atividade. O estatuto associado a um determinado indivíduo pode ser mecânico, não mecânico, nobre ou ser desconhecido. O estatuto é Mecânico ou Não Mecânico consoante o trabalho dos indivíduos seja, ou não, manual. Mecânico corresponde ao oposto de nobre. Remete para aquele que trabalha com as mãos e que tem de trabalhar para sobreviver. Deste modo está "mais abaixo" na hierarquia social um indivíduo cujo estatuto é mecânico. Não Mecânico é aquele que não é nobre, mas também não é exatamente mecânico ou quando não se tem a certeza que é nobre. O estatuto Nobre corresponde a todos os indivíduos que tenham um título de nobreza. Sempre que o estatuto é desconhecido considerou-se que o estatuto era Não Se Sabe.

A partir da Tabela 5.6 podem-se obter os estatutos por setor de atividade (Figura 5.2), bem como os estatutos por áreas de ocupação (Figura 5.3). Salienta-se o facto de apenas três setores de atividade (Administração, Doméstico, Ensino) apresentarem ocorrências de todos os estatutos e quatro setores de atividade terem somente um estatuto (Pesca, Pecuária, Mineração, Diplomacia). Relativamente às ocupações, três apresentam indivíduos dos quatro tipos de estatuto (Governo da Res Publica, Governo da Casa, Artes Liberais) e três apresentam apenas um só estatuto (Pesca, Mineração, Embaixadas).

Tabela 5.6: Tipo de estatuto por ocupação/setor de atividade.

Ocupação (designação da época)	Sector de atividade	Estatuto			
Artes liberais	Administração		NM		
	Ensino		NM		
	Justiça		NM	No	
	Saúde		NM		NS
Artes mecânicas	Ensino	Me			NS
	Saúde	Me	NM		NS
	Transformação	Me	NM		
	Transporte	Me			
Belas Artes	Belas Artes	Me	NM		NS
Embaixadas	Diplomacia			No	
Estudo	Ensino	Me	NM	No	
Governo da casa	Administração	Me	NM	No	
	Doméstico	Me	NM	No	NS
	Não se sabe		NM		
Governo da res publica	Administração	Me	NM	No	NS
	Defesa		NM		
	Eclesiástico		NM		
	Justiça	Me	NM	No	
	Saúde		NM		
	Segurança	Me	NM	No	
	Transporte	Me	NM		
	Igreja	Eclesiástico		NM	No
Lavoura	Agricultura	Me	NM	No	
	Pecuária	Me			
Mercancia	Negócio	Me	NM		
Milícia	Defesa	Me	NM	No	
	Segurança	Me	NM	No	
Mineração	Mineração		NM		
Não se sabe	Administração	Me	NM		NS
	Não se sabe	Me	NM		
Navegação	Navegação	Me	NM	No	
Pesca	Pesca	Me			
Sem ocupação	Sem ocupação		NM	No	
Sem ocupação/nobre	Sem ocupação/nobre		NM	No	
Viver da sua fazenda	Viver da sua fazenda		NM	No	

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe

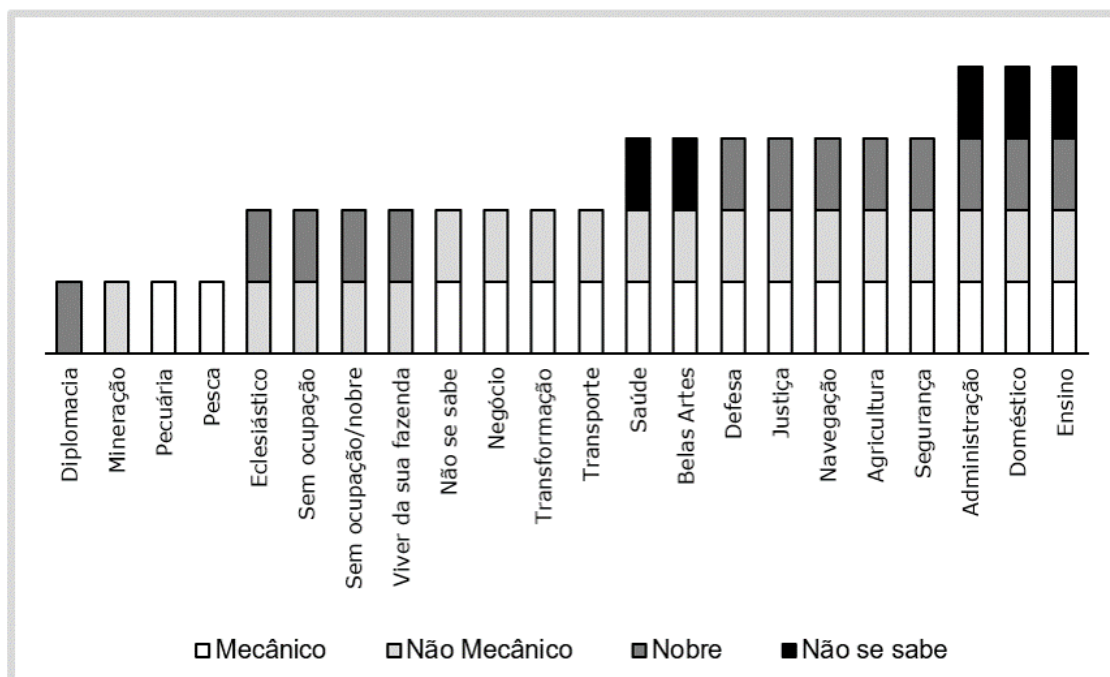


Figura 5.2: Distribuição dos estatutos por setor de atividade.

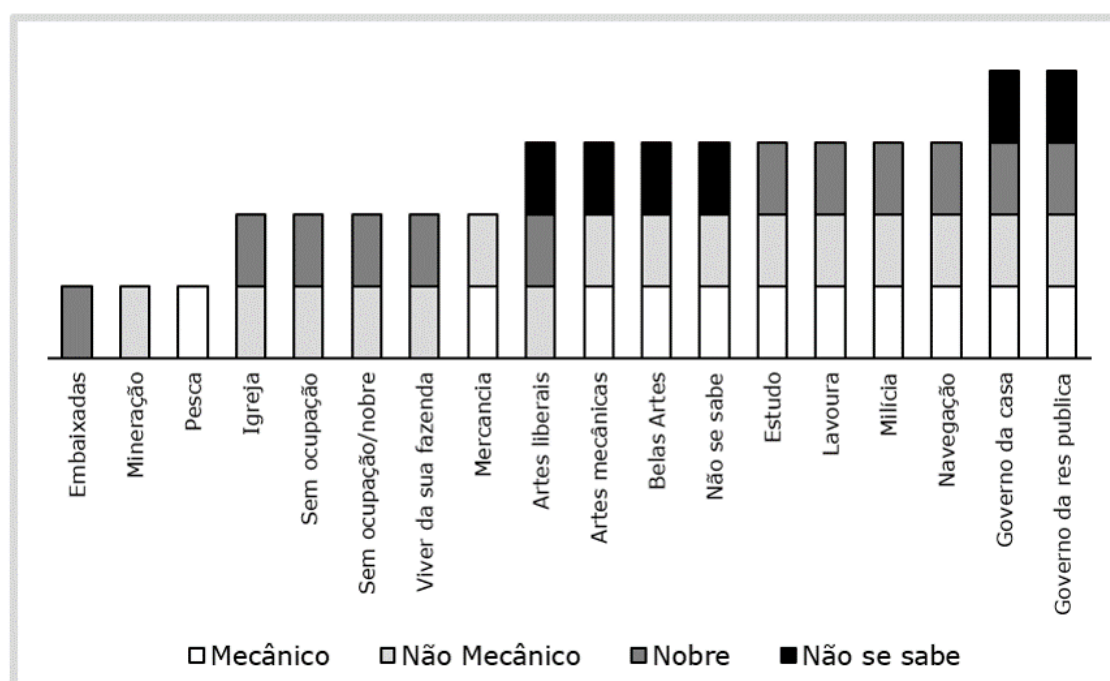


Figura 5.3: Distribuição dos estatutos por áreas de ocupação na designação da época.



O domínio que as ontologias representam é o de múltiplas ocupações de egos identificados quase sempre por ligações ao SO. Essas profissões e cargos foram classificados em duas grelhas principais: estatuto da ocupação na época e setor de atividade.

A aplicação dessa grelha foi feita por historiadores, que tiveram em linha de conta vários parâmetros sobre o indivíduo em causa e não apenas uma simples informação sobre a ocupação de uma dada personagem.

Para a classificação foram essenciais os dados de contexto. Note-se que na realidade classificaram-se pessoas e não isoladamente ocupações. Estas últimas em si mesmas podem admitir rótulos por vezes opostos. Cite-se o caso de médico ou de lavrador. Quando um médico equivalia a alguém com formação universitária, o cargo foi classificado como Arte Liberal e era Não Mecânico; quando não foi classificado como Arte Liberal, era Mecânico. Lavrador de terras próprias foi incluído como Não Mecânico no século XVIII e entre as Artes Mecânicas quando o lavrador tinha de trabalhar ainda em terras alheias por jornal para poder sobreviver. Na realidade, o que se pretende é classificar automaticamente indivíduos, em bases de dados prosopográficas, sem fugir à complexidade da época considerada.

## 5.3 Identificação de palavras-chave

Com o objetivo de classificar automaticamente ocupações em dados pouco estruturados, realizou-se um levantamento das palavras-chave que poderiam ser utilizadas para a referida classificação. Nesse sentido foi necessário realizar diversas operações, nomeadamente, levantamento de valores únicos e triagem de palavras-chave. Ao longo de todo este percurso foi realizado ainda um levantamento e correção de erros ortográficos.

### 5.3.1 Levantamento de valores únicos

Identificaram-se 23482 registos referentes a ocupações. Sobre estas ocorrências efetuaram-se pesquisas, de modo a obter apenas os valores únicos. Assim, por exemplo, a ocupação só com a palavra "Lavrador" acontece em 55 registos. Agregou-se todas essas ocorrências numa única. Após efetuar este tipo de agrupamento em todas as ocupações, obtiveram-se 8426 registos (27 de agosto de 2015). Foi esta a base de trabalho sobre a qual se realizou a triagem de palavras-chave.

### 5.3.2 Triagem de palavras-chave

Sobre todas as ocorrências únicas determinou-se qual, ou quais, a(s) palavra(s) que serão utilizadas para a classificação dos setores de ocupação, dos setores de atividade e dos estatutos. Em muitas situações, uma só palavra era insuficiente. Apresentamos na Tabela 5.7 o exemplo para a palavra "Administrador". Nesta situação, a escolha de apenas uma palavra (P1 na Tabela 5.7) não chegava para classificar setor de ocupação, setor de atividade ou estatuto.

Tabela 5.7: Exemplo de triagem para a palavra Administrador.

P1	P2	Setor de ocupação	Setor de atividade	Estatuto
Administrador	Escravos	Mercancia	Negócio	Mecânico
Administrador	Fazendas	Lavoura	Agricultura	Não Mecânico
Administrador	Sabão	Mercancia	Negócio	Não Mecânico

No Apêndice B podem consultar-se as palavras-chave encontradas. Para obter esta listagem foi necessário analisar:

- dicionários da época<sup>3</sup> e arquivos digitais<sup>4</sup>, consultados durante o mês de agosto de 2015;
- o contexto da frase;
- o leque ocupacional do indivíduo;
- a triagem por uma equipa de especialistas.

Por vezes ainda foi necessário ter presente o ano da ocorrência.

A preocupação pela classificação de ocupações históricas originou a criação da *Historical International Standard Classification of Occupation*<sup>5</sup> (HISCO). A criação desta classificação normalizada resultou da necessidade de incorporar em classificações internacionais como a ISCO as ocupações brasileiras a partir do século XIX.

Tratando-se de um projeto que teve a primeira versão em 1958, a ISCO utilizada como base de trabalho para a construção da HISCO foi a de 1968. Botelho, van Leeuwen, Maas e Miles (2006) referem que não optaram pela versão mais recente pois esta "perdeu" algumas das ocupações históricas presentes nas primeiras versões.

Uma das razões porque as ocupações da HISCO não foram escolhidas, como base da classificação, foi o passado recente que representam. Ao considerarem que "a ocupação é uma variável chave em muitos campos da história", Botelho et al. (2006, p. 168) sentiram a necessidade de concretizar este projeto. Assim, a designação atribuída a uma dada ocupação entra em linha de conta com duas condicionantes principais, nomeadamente período temporal e espaço geográfico. Variáveis estas que são também fundamentais na classificação dos eventos históricos portugueses, com uma grande diferença: no SPARES importa recuar aos séculos XVI-XVIII.

Considerou-se nalguns casos que a classificação dependia do contexto da ocupação (última coluna das tabelas visualizadas no Apêndice B). Estas situações podem ser consultadas na Tabela 5.8.

Pela observação da Tabela 5.8 destaca-se:

- "cirurgião" é a única situação em que o setor de ocupação e o estatuto dependem ambos do contexto;
- o setor de ocupação com maior número de ocorrências a depender do contexto é "Governo da res pública";
- o setor de atividade com maior número de ocorrências a depender do contexto é "Administração";
- "Segurança", "Negócio" e "Justiça" apresentam apenas uma ocorrência para o setor de atividade;
- "Milícia" aparece apenas um vez a depender do contexto como setor de ocupação.

Pela análise das palavras-chave encontradas pode constatar-se que surgiram algumas situações que não estavam previstas na classificação inicial, referida na Figura 5.1. Estas novas ocorrências foram classificadas relativamente ao setor de ocupação, ao setor de atividade e ao estatuto, como se pode observar na Figura 5.4.

---

<sup>3</sup>por exemplo <http://www.brasiliana.usp.br/en/dicionario>

<sup>4</sup>por exemplo <http://digitarq.adevr.arquivos.pt>

<sup>5</sup><http://www.ilo.org/public/english/bureau/stat/isco>

Tabela 5.8: Palavras-chave com classificação a depender do contexto.

Palavras-chave	Setor de ocupação	Setor de atividade	Estatuto			
			ME	NM	NO	NS
adjutório	Governo da casa	Transformação	x			
cirurgião		Saúde				
depositário	Governo da casa	Segurança				
lavrador	Lavoura	Agricultura	x			
lavradora	Lavoura	Agricultura	x			
médico	Não se sabe	Saúde				x
médico; assistente	Artes liberais	Saúde		x		
membro; junta; estados	Governo da res publica	Administração		x		
merceeira	Não se sabe	Não se sabe	x			
mestre; estudantes	Artes liberais	Ensino		x		
mestre; gramática	Artes liberais	Ensino		x		
moço; saco	Não se sabe	Não se sabe	x			
músico	Belas artes	Belas artes				x
não; tem; ocupação	Sem ocupação	Sem ocupação				x
não; tem; ofício	Sem ocupação	Sem ocupação				x
oficial; justiça	Governo da res publica	Justiça		x		
organista	Belas artes	Belas artes				x
pintor	Artes mecânicas	Transformação	x			
pintor; louça	Artes mecânicas	Transformação	x			
prioste	Mercancia	Negócio	x			
procurador	Governo da casa	Administração				
procurador; concelho	Governo da res publica	Administração				
recondução; serventia	Governo da res publica	Administração	x			
secretário; câmara	Governo da res publica	Administração				
sem; ocupação	Sem ocupação	Sem ocupação				
sem; ofício	Não se sabe	Não se sabe				
serve; majestade	Governo da res publica	Não se sabe		x		
serve; escrivão	Governo da res publica	Administração				
serve; marinha	Navegação	Defesa		x		
serve; tesoureiro	Governo da res publica	Administração				
serviu; casa; beneficiado	Governo da res publica	Doméstico				
serviu; conde	Governo da casa	Doméstico				
serviu; rei	Milícia	Defesa		x		
sub-tesoureiro	Governo da res publica	Administração	x			
tesoureiro	Governo da res publica	Administração				
tutor	Governo da casa	Administração		x		
vive; lavoura	Lavoura	Agricultura	x			
vive; rendimento; letras	Artes liberais	Viver da sua fazenda		x		

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe

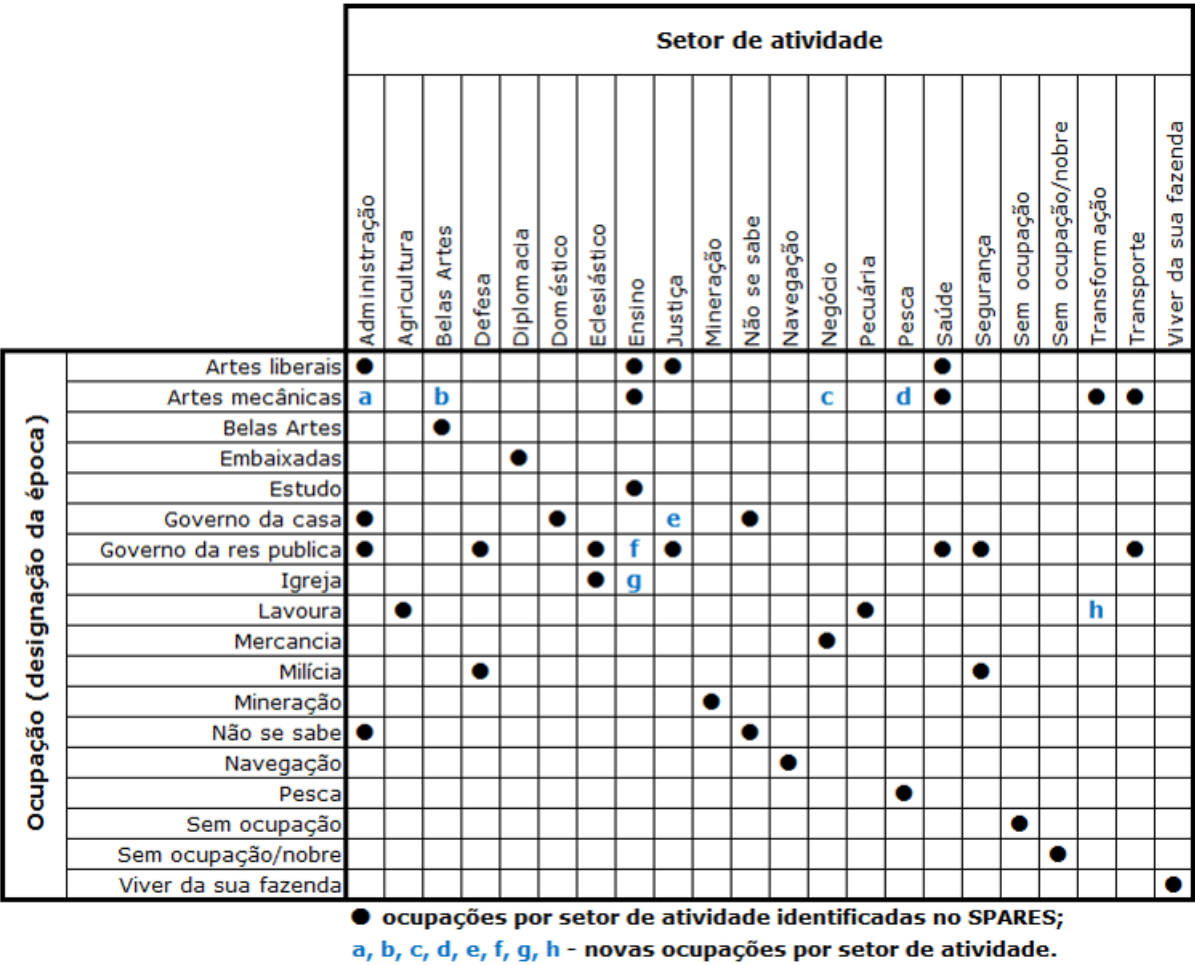


Figura 5.4: Distribuição dos estatutos por setor de ocupação, após triagem de palavras-chaves.

Às novas ocorrências, identificadas na Figura 5.4, apontam-se como exemplo os ofícios que podem consultar-se na Tabela 5.9.

Tabela 5.9: Exemplos de novos ofícios.

a	selador de alfândega	b	chameleiro
c	aguadeiro	d	armador de sardinha
e	curador	f	reformador numa universidade
g	leitor	h	proprietário de engenho de açúcar

5.4 Extração de eventos

Tratando-se de um repositório central construído em MySQL, o SPARES torna possível a utilização de interfaces que permitem o seu preenchimento. Na Figura 5.5 pode observar-se o exemplo parcial de uma das interfaces que se utilizam. O texto extraído tem de ser preparado para que o resultado possa ser confrontado com as palavras-chave constantes do Apêndice B.

Tipo de Evento	Evento	Localidade
Relacional	<Agente institucional>Comissário SO	2
Biográfico	<Formação Acadêmica>	1
Biográfico	<Ocupação>Vigário confirmado da	8
Biográfico	<Ocupação>Cónego da Sé do	6

Figura 5.5: Apresentação parcial de uma interface.

Para que uma nova ocupação possa ser atribuída no campo Evento utiliza-se uma interface cuja abertura é acionada por duplo click nesse campo. Como resultado surge a interface apresentada na Figura 5.6, que na imagem representa a inserção de um evento com a ocupação de "Alfaiate".

Figura 5.6: Interface para preenchimento de eventos.

Esta interface tem na sua construção e preenchimento a utilização de etiquetas em *Extensible Markup Language* (XML). Como resultado da inserção visualizada na Figura 5.6, no repositório SPARES o campo Evento é preenchido como se observa na Figura 5.7.

Evento
<div>&lt;Ocupação&gt;Alfaiate&lt;/Ocupação&gt;&lt;Observações&gt;Tem estabelecimento próprio&lt;/Observações&gt;&lt;/div>

Figura 5.7: Exemplo de evento preenchido na tabela Comissarios.

Para que se possa concretizar o processamento em linguagem natural, é necessário identificar todos os caracteres que constituem as ocorrências de XML no campo Evento. Posteriormente, estas serão substituídas para que este campo possa ser devidamente interpretado.

Foram, assim, identificadas as ocorrências relacionadas com o início, ou fim, de uma etiqueta em XML, bem como os caracteres que as irão substituir (Tabela 5.10).

Tabela 5.10: Ocorrências de XML e caracter de substituição.

Ocorrência de XML:	Substituir por:
<div>&lt;	<
&gt;	>
&lt;	<
&gt;</div>	>

Como resultado da aplicação das substituições mencionadas na Tabela 5.10, o evento anteriormente visualizado na Figura 5.7 passa a ter a seguinte constituição:

**<Ocupação>Alfaiate</Ocupação><Observações>Tem estabelecimento próprio</Observações>**

Estas alterações refletir-se-ão em qualquer evento. Como o objetivo é extrair apenas o evento ocupação, conjugaram-se na extração as etiquetas XML com a palavra Ocupação (Tabela 5.11).

Tabela 5.11: Ocorrências de XML e respetiva substituição.

Ocorrência de XML:	Substituir por:
<div>&lt;Ocupação&gt;	<Ocupação>
&lt;Ocupação&gt;	</Ocupação>

Constatou-se que estas ocorrências só aconteciam quando a ocupação não era o único evento inserido. Nas situações em que ocupação é o único evento, os dados inseridos já são do tipo:

**<Ocupação>Alfaiate</Ocupação>**

Para além das já referidas, ocorreram ainda outras situações, resultantes da aplicação de formatações em XML (fonte, tamanho ou estilo da letra). Para as ocorrências que se apresentam na Tabela 5.12, a substituição aplicada foi o valor NULL.

Tabela 5.12: Ocorrências de XML substituídas por valor NULL.

Ocorrência de XML:	Substituir por:
&nbsp;	NULL
&quot;	
<div>	
</div>	

Após identificação destas ocorrências efetuaram-se os procedimentos necessários para obter o texto que irá ser utilizado na aplicação das duas técnicas de PLN. Foram criadas duas pesquisas em *Structured Query Language* (SQL), que podem ser consultadas nas figuras 5.8 e 5.9.

```
SELECT Evento, Replace(Evento,"<div>&lt;Ocupação&gt;","<Ocupação>") AS Part1,
Replace(Evento,"&lt;Ocupação&gt;","</Ocupação>") AS Part2 FROM comissarios
WHERE (Evento Like "*Ocupação*") ORDER BY Evento;
```

Figura 5.8: Pesquisa que aplica as substituições referidas na Tabela 5.11.

```
SELECT Part2, Replace(Part2,"<div>","") AS Part3, Replace(Part3,"</div>","") AS Part4,
Replace(Part4,"&nbsp;","") AS Part5, Replace(Part5,"&quot;","") AS EventoOcup
FROM 1ExtrairOcup WHERE Part2 Like "<Ocupação*" And Part2 Like "*Ocupação>"
ORDER BY Part2;
```

Figura 5.9: Pesquisa que aplica as substituições referidas na Tabela 5.12.

O resultado destas duas pesquisas permitiu obter os eventos, como se exemplifica na Figura 5.10. A Figura 5.11 apresenta a pesquisa que irá transformar estes eventos de ocupação no texto que posteriormente será processado. O resultado da pesquisa representada na Figura 5.11, aplicado aos registos visualizados na Figura 5.10, permite obter o resultado que se apresenta na Figura 5.12.

EventoOcup
<Ocupação>Vivia de suas rendas e quintas</Ocupação>
<Ocupação>Volanteiro</Ocupação>
<Ocupação>Volanteiro</Ocupação>
<Ocupação>Vivia de sua fazenda</Ocupação>
<Ocupação>Vivia de sua fazendas e rendas que tomava</Ocupação>
<Ocupação>Vivia de suas fazendas e vinhas</Ocupação>
<Ocupação>Vivia de suas fazendas</Ocupação>

Figura 5.10: Exemplo das ocupações após execução das pesquisas apresentadas nas figuras 5.8 e 5.9.

```
SELECT Replace(EventoOcup,"<Ocupação>","") AS Part6,
Replace(Part6,"</Ocupação>","") AS Eventos FROM 2ExtrairOcup;
```

Figura 5.11: Pesquisa que transforma os eventos de ocupação em texto que será processado por ferramentas de PLN.

Eventos
Vivia de suas rendas e quintas
Volanteiro
Volanteiro
Vivia de sua fazenda
Vivia de sua fazendas e rendas que tomava
Vivia de suas fazendas e vinhas
Vivia de suas fazendas

Figura 5.12: Exemplo das ocupações após execução da pesquisa apresentada na Figura 5.11.

A remoção de *stopwords* e aplicação de *stemming* realiza-se após esta preparação dos dados aos eventos obtidos.

## 5.5 Remoção de *stopwords* e aplicação de *stemming*

Uma das hipóteses a testar é saber se o PLN pode contribuir para uma mais rápida e eficaz classificação dos eventos ocorridos entre dois quaisquer intervenientes do SO. Para tal, e após a determinação manual das palavras-chave necessárias à classificação, realizou-se um ensaio com os eventos que continham a palavra VIVE (num total de 159 registos). As palavras-chave encontradas por evento são constituídas no mínimo por 2 palavras (por exemplo: vive|agência) e no máximo por 4 palavras (por exemplo: vive|ordens|bens|patrimoniais). Deste modo, sempre que num evento ocorram simultaneamente as palavras-chave vive e agência este deverá ser classificado como: ocupação – mercancia; setor de atividade – negócio; estatuto – não mecânico. Por outro lado, se as palavras-chave forem vive, ordens, bens e patrimoniais, o evento deve ser classificado como: ocupação – igreja; setor de atividade – eclesiástico; estatuto – não mecânico.

Para a preparação dos dados prosopográficos históricos, necessária para concretizar a classificação dos eventos, foram fundamentais algumas técnicas de PLN, nomeadamente: i) remoção de *stopwords*; ii) aplicação de *stemming*. As *stopwords* são palavras irrelevantes no âmbito da classificação dos dados realizada. A aplicação do método de *stemming* permite reduzir as palavras ao seu radical. Na Tabela 5.13 apresenta-se um evento conforme ele é escrito na base de dados SPARES e o resultado após a remoção de *stopwords* e aplicação do método de *stemming*.

Tabela 5.13: Visualização de evento sem *stopwords* e após *stemming*.

Evento em SPARES	vive de sua agência e contrato de sola e gados, e dos ordenados que tem de procurador
Após remoção de <i>stopwords</i>	vive agência contrato sola gados ordenados procurador
Após aplicação de <i>stemming</i>	viv agênc contrat sol gad orden procur

Para se poder confrontar os resultados obtidos na classificação manual de eventos com os resultados provenientes do PLN, extraiu-se da base de dados SPARES os eventos que começavam com a palavra **vive**, um dos agrupamentos ocupacionais mais complexos para os historiadores, não obstante a sua ocorrência frequente. Este conjunto admite um amplo padrão de variabilidade: vive de sua fazenda, vive da sua agência e trabalho; vive das suas letras; vive das suas ordens, etc. Criou-se, com estes, uma nova tabela. Para tal recorreu-se à instrução visualizada na Figura 5.13.

```
SELECT Evento FROM EventosTAB INTO EventosVIVE
WHERE Evento LIKE "vive *" ORDER BY Evento;
```

Figura 5.13: Criação da tabela Evento.

Aos registos deste modo obtidos, ir-se-ão retirar as palavras consideradas irrelevantes no contexto do evento – *stopwords*. Para a extração destas palavras consultaram-se, em 4 de fevereiro de 2016, diversas listagens de *stopwords*. Apresenta-se na Tabela 5.14 as localizações das listagens de *stopwords* que foram utilizadas para a lista utilizada neste trabalho.

Para além das listagens atrás referidas consultou-se ainda a lista de *stopwords* que a aplicação PostgreSQL disponibiliza.

Constatou-se que algumas das *stopwords* eram palavras-chave, anteriormente classificadas manualmente, tendo assim de ser removidas. Apresenta-se na Tabela 5.15 as palavras que foram retiradas da lista inicial de *stopwords* por serem relevantes para a classificação que foi realizada.



Tabela 5.14: Localização das listas de *stopwords*.

---

<http://members.unine.ch/jacques.savoy/clef/index.html>  
<https://code.google.com/archive/p/stop-words>  
<http://snowball.tartarus.org/algorithms/portuguese/stop.txt>  
<https://gist.github.com/alopes/5358189>  
<http://miningtext.blogspot.pt/2008/11/listas-de-stopwords-stoplist-portugues.html>  
<http://www.ranks.nl/stopwords/portuguese>

---

Tabela 5.15: *Stopwords* removidas da lista inicial.

algum	fazer	meio	primeiro	terceiro
ao	fazia	menor	relação	trabalhar
conselho	feitos	na	sem	trabalho
da	fora	não	sete	usa
do	geral	número	sua	vários
em	grande	para	suas	
estado	ligado	povo	tem	
faz	maio	primeira	terceira	

Observou-se ainda que havia palavras que surgiam repetidamente e cujo contexto não é relevante. Estas foram introduzidas na lista de *stopwords* (Tabela 5.16).

Tabela 5.16: Palavras introduzidas na lista de *stopwords*.

ágil	estrangeiras	paroquial	santa
bairro	freguesia	penitenciária	são
distrito	irmandade	s.	

Para além das palavras referidas na Tabela 5.16 muitos dos eventos possuem no seu conteúdo nomes de freguesias, concelhos ou comarcas. Estes nomes consideraram-se irrelevantes para a classificação (Tabela 5.17).

Tabela 5.17: Nomes de freguesias, concelhos ou comarcas irrelevantes para a classificação.

Arcos de Valdevez	Felgueiras	Porto	Viana
Azurar da Beira	Freixada	Quintela	Viana da Foz do Lima
Braga	Funchal	Resende	Vila Real
Canas de Senhorim	Gondomar	Sanfins	Viseu
Caniçada	Lamego	Soutelo	
Covas de Barroso	Maçal da Ribeira	Unhão	
Douro	Moncorvo	Vandoma	

Os nomes constantes da Tabela 5.17 deram origem às *stopwords* que se apresentam na Tabela 5.18.

A lista final de *stopwords* (constituída por 648 *stopwords*) foi assim o resultado das várias listas de *stopwords* consultadas, da remoção das palavras referidas na Tabela 5.15 e da introdução das palavras referidas nas tabelas 5.16 e 5.18. É, digamos, uma lista especializada em função da temática em análise: ocupações.

Tabela 5.18: *Stopwords* resultantes dos nomes de freguesias, concelhos ou comarcas.

arcos	braga	douro	funchal	maçal	real	senhorim	vandoma
azurar	canas	felgueiras	gondomar	moncorvo	resende	soutelo	viana
barroso	caniçada	foz	lamego	porto	ribeira	unhão	vila
beira	covas	freixada	lima	quintela	sanfins	valdevez	viseu

Substituiu-se a lista que o PostgreSQL disponibiliza pela lista de *stopwords* que pode ser consultada no Apêndice C. Após esta operação construíram-se as instruções, presentes na Figura 5.14, em SQL.

```
ALTER TABLE "EventosVIVE" ADD COLUMN TextoVetor TSVECTOR;
UPDATE "EventosVIVE" SET TextoVetor = TO_TSVECTOR("Evento");
```

Figura 5.14: Criação e preenchimento do campo TextoVetor.

Como resultado da execução destas instruções é adicionada uma nova coluna à tabela EventosVIVE, denominada TextoVetor. A aplicação registou para este campo as palavras reduzidas ao seu radical, após remoção das *stopwords*, seguido da posição dessas palavras na frase. Por exemplo 'fazend':4 é o resultado de um evento cuja 4ª palavra da frase é fazenda ou fazendas.

Verificou-se assim que, para eventos diferentes do campo Evento, podem ocorrer correspondências iguais em TextoVetor. Na Tabela 5.19 podem ser consultados eventos que estão na origem dos radicais **fazend** e **viv**.

Tabela 5.19: Exemplo de eventos com a mesma correspondência no campo TextoVetor.

Evento	TextoVetor
Vive também de sua fazenda	'fazend':5 'viv':1
Vive das duas fazendas	'fazend':4 'viv':1
Vive das suas fazendas	'fazend':4 'viv':1
Vive de sua fazenda	'fazend':4 'viv':1
Vive de suas fazendas	'fazend':4 'viv':1
Vive por sua fazenda	'fazend':4 'viv':1

Das palavras-chave que fazem parte do Apêndice D, consideraram-se apenas as que correspondem aos eventos que foram preenchidos com as palavras-chave obtidas pelo processo manual. Resultaram deste modo 100 palavras-chave distribuídas por 9 categorias (número máximo de palavras-chave encontradas num evento).

## 5.6 Classificação de eventos ...

### 5.6.1 ... a partir das palavras-chave obtidas manualmente

Os eventos foram classificados, relativamente à ocupação, atividade e estatuto, através da execução de 4 pesquisas (Figura 5.15). Pode observar-se na Tabela 5.20, como exemplo, a classificação de 4 eventos após a execução das instruções referenciadas na Figura 5.15.

Após esta classificação, apenas 117 dos 159 eventos extraídos, referidos na secção 5.4, tiveram correspondência com as palavras-chave encontradas manualmente.

```
SELECT DISTINCT Evento, P1, P2, P3, P4, InStr(1,Evento,[P1]) AS a, InStr(1,Evento,P2) AS b,
InStr(1,evento,P3) AS c, InStr(1,Evento,P4) AS d, ocup, ativ, ME, NM, PC_ClassifAGRUP.NO, NS,
dep FROM EventosVIVE117, PC_ClassifAGRUP;
```

```
SELECT DISTINCT IIf(a>0,1,0) AS aa, IIf(b>0,1,0) AS bb, IIf(c>0,1,0) AS cc, IIf(d>0,1,0) AS dd,
aa+bb+cc+dd AS ee, Evento, ocup, ativ, ME, NM, Cvive1.[NO] AS NOb, NS, dep FROM Cvive1;
```

```
SELECT Max(ee) AS Maximo, Evento FROM Cvive2 GROUP BY Evento;
```

```
SELECT DISTINCT Cvive3.Maximo, IIf(maximo=ee, cvive3.Evento) AS EventoF, ocup, ativ, ME,
NM, NOb, NS, dep FROM Cvive2 INNER JOIN Cvive3 ON Cvive2.Evento = Cvive3.Evento
WHERE (IIf(maximo=ee,cvive3.Evento) Is Not Null);
```

Figura 5.15: Pesquisas utilizadas para a classificação dos eventos.

Tabela 5.20: Exemplo da classificação para 4 eventos.

Evento	P1	P2	P3	Ocupação	Atividade	Estatuto
Vive da sua missa	vive	missa		Igreja	Eclesiástico	NM
Vive de algum contrato	vive	contrato		Mercancia	Negócio	NM
Vive de pano de linho	vive	pano	linho	Artes mecânicas	Transformação	ME
Vive do seu trabalho e agência	vive	agência	trabalho	Lavoura	Agricultura	ME

### 5.6.2 ... a partir das palavras-chave obtidas por remoção de *stopwords* e aplicação de *stemming*

Contrariamente ao que aconteceu com a classificação a partir das palavras-chave obtidas manualmente, a extração de radicais nos 159 eventos está diretamente associada ao evento mas não tem ocupação, atividade ou estatuto associado.

Para se poder comparar os dois métodos aplicados, é necessário confrontar as palavras-chave através deles obtidas. Assim, realizou-se a remoção de *stopwords* e a aplicação de *stemming* às palavras-chave obtidas de modo manual. Estas foram comparadas com os radicais obtidos diretamente sobre os eventos.

Foram identificadas duas situações em que as palavras-chave obtidas manualmente eram *stopwords*, nomeadamente "sem" e "suas". Estas encontravam-se nos seguintes grupos de palavras-chave: i) vive|fazenda|sem|ofício; ii) vive|suas|terras. Na nossa amostra estas ocorrências correspondem à classificação de dois eventos na primeira situação e um evento na segunda.

Para atribuir uma ocupação, uma atividade e um estatuto aos eventos cujas palavras-chave foram obtidas por remoção de *stopwords* e aplicação de *stemming*, estes foram comparados com os eventos classificados manualmente. Para tal executaram-se as instruções que se apresentam na Figura 5.16.

```
SELECT PC_Manual_S.Evento, If(PC_Manual_S.Evento=PC_ClassifAGRUPS.Evento,Ocup)
AS OcupC, If(PC_Manual_S.Evento=PC_ClassifAGRUPS.Evento,Ativ) AS AtivC,
If(PC_Manual_S.Evento=PC_ClassifAGRUPS.Evento,ME) AS MEC,
If(PC_Manual_S.Evento=PC_ClassifAGRUPS.Evento,NM) AS NMC,
If(PC_Manual_S.Evento=PC_ClassifAGRUPS.Evento,[NO]) AS NOC,
If(PC_Manual_S.Evento=PC_ClassifAGRUPS.Evento,[NS]) AS NSC,
If(PC_Manual_S.Evento=PC_ClassifAGRUPS.Evento,Dep) AS DepC
INTO PC_ManualClass FROM PC_Manual_S INNER JOIN PC_ClassifAGRUPS ON
PC_Manual_S.Evento = PC_ClassifAGRUPS.Evento;
```

Figura 5.16: Instruções utilizadas para a atribuição de ocupação, atividade e estatuto aos eventos cujas palavras-chave foram obtidas automaticamente.

## 5.7 Criação de cenários ontológicos

Tendo por base os conhecimentos adquiridos com a classificação dos eventos, foi criado um suporte formal para uma aproximação integrada à prosopografia que pudesse complementar a análise tradicional dos dados sociais. Nesse sentido foram desenvolvidos dois cenários ontológicos.

Para a visualização da hierarquia das classes construídas, utilizou-se o *plug-in* OntoGraf<sup>6</sup> e para visualização dos restantes elementos o *plug-in* SOVA<sup>7</sup>. Através destes *plug-ins* é ainda possível exportar imagens dessas visualizações.

A classe que contém todos os indivíduos, ou seja, todas as subclasses, designa-se por *owl:Thing* (Horridge, 2011). Na definição dos nomes a atribuir às classes, seguiu-se a sugestão de Horridge (2011): começam todos por letra maiúscula e não contém espaços. Quando foi necessário mais do que um nome para a mesma classe, juntaram-se os nomes (por exemplo: ArtesLiberais). Sempre que um nome era uma palavra acentuada retiraram-se os acentos e substituiu-se a letra ç pela letra c (por exemplo: Ocupacao). Sempre que duas classes possuíam subclasses com nomes iguais colocou-se, no término do nome, At (para setor de atividade), Oc (para área de ocupação na designação da época) ou Est (para estatuto). Por exemplo: PescaAt, PescaOc.

Considerou-se, na construção da hierarquia de classes, que todas as subclasses criadas são classes disjuntas, ou seja, um indivíduo que pertença a uma das classes referidas não pode pertencer a nenhuma das outras (Horridge, 2011).

À semelhança do que aconteceu nos nomes atribuídos às classes, também para as propriedades de objeto e para as propriedades de tipos de dados se seguiu a sugestão de Horridge (2011): os nomes das propriedades começam por letra minúscula, com o prefixo *tem* ou *é*; quando compostas por mais de uma palavra não devem conter espaços e a partir da segunda palavra começam por letra maiúscula. Uma propriedade de objeto relaciona indivíduos de um domínio com indivíduos de um contradomínio. Uma propriedade de tipo de dados relaciona indivíduos de um domínio a tipo de dados de um contradomínio.

A todas as propriedades de objeto começadas com o prefixo *tem* foi atribuído domínio e contradomínio. Para todas estas propriedades de objeto foram criadas as propriedades inversas correspondentes, com o prefixo *é*. Deste modo, por exemplo, quando se criou a propriedade *temMecanico* criou-se também a propriedade inversa *éMecanicoDe*. Não é necessário atribuir-lhe um domínio e um contradomínio, uma vez que ao ser definida como inversa de *temMecanico* assume automaticamente como domínio o contradomínio desta e como

<sup>6</sup><http://protegewiki.stanford.edu/wiki/OntoGraf>

<sup>7</sup><http://protegewiki.stanford.edu/wiki/SOVA>

contradomínio o seu domínio.

O povoamento da ontologia foi feito, nos dois cenários, com recurso aos dados disponibilizados no SPARES. Para tal efeito, usou-se o *software* Ontop<sup>8</sup> disponibilizado gratuitamente e desenvolvido pela Universidade de Bozen-Bolzano.

Para utilizar o *plug-in* Ontop foi necessário realizar alguns procedimentos. Atualmente a última versão de Protégé já tem associado este *plug-in*. É, no entanto, necessário colocar dois separadores no ambiente da aplicação (*ontop Mappings* e *ontop SPARQL*). Com o separador *ontop Mappings* é possível gerir o acesso a bases de dados externas e construir o modelo OBDA (*Ontology Based Data Access*). Este separador divide-se ainda em três janelas distintas: *Datasource manager*; *Mapping manager*; *Mapping Assistant – BETA*.

A primeira operação realizada foi a conexão à base de dados. As tabelas utilizadas encontram-se disponíveis no repositório de dados.

Equacionaram-se duas abordagens diferentes para a construção das ontologias.

Num primeiro cenário a ontologia foi criada automaticamente por transposição do sistema SPARES. Para tal utilizaram-se as ferramentas disponibilizadas pelo *plug-in* Ontop: i) *Generate ontology and mappings...*; ii) *Materialize triples*. A primeira cria a estrutura da ontologia à semelhança do sistema SPARES, tabelas e campos, respetivamente classes e propriedades de tipo de dados no modelo ontológico. A segunda ferramenta adiciona instâncias à ontologia (dados constantes nos campos das tabelas).

O segundo cenário integrou quatro etapas de desenvolvimento. Numa primeira fase construíram-se as classes. De seguida definiram-se as propriedades de objeto que permitem relacionar as classes criadas. As propriedades de tipo de dados foram construídas na terceira etapa. Por último, adicionaram-se instâncias. As três primeiras etapas foram realizadas utilizando as ferramentas disponibilizadas no Protégé. Para realizar a quarta etapa foi necessário o preenchimento automático das instâncias. Para tal criaram-se os *mappings* necessários, utilizando o separador *Mapping Manager*.

Exemplifica-se, com as instruções presentes na Figura 5.17, os *mappings* necessários para o povoamento da classe Localizacoes.



Figura 5.17: *Mappings* construídos para povoar a classe Localizacoes.

As instruções referidas na Figura 5.17 tiveram de ser materializadas através da ferramenta *Materialize triples*.

Sobre as ontologias podem ser realizadas diversas interrogações. Estas exemplificam algumas das pesquisas tipo que podem ser aplicadas a estes dados e na sua construção utilizou-se a ferramenta *SPARQL Query* que a aplicação Protégé disponibiliza. Quando se inicializa pela primeira vez, esta ferramenta apresenta as instruções que podem ser consultadas na Figura 5.18.

<sup>8</sup><http://ontop.inf.unibz.it>

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT ?subject ?object
WHERE { ?subject rdfs:subClassOf ?object }
```

Figura 5.18: SPARQL Query: instruções iniciais.

A execução das instruções referidas na Figura 5.18 terá como resultado a estrutura hierárquica de classes que a ontologia apresenta.

Para além dos prefixos que por defeito são colocados inicialmente pela ferramenta SPARQL Query, cuja declaração permite a identificação de recursos que podem ser utilizados, também é necessário adicionar um prefixo que identifique a ontologia que vai ser interrogada. Observa-se na Figura 5.19 o prefixo adicionado.

```
PREFIX : <http://OntoSpares/Rev/2016/7/2/OntoSPARES#>
```

Figura 5.19: Prefixo declarado para OntoSPARES.

Nas consultas realizadas, que se apresentam na secção 6.4, para além dos prefixos anteriormente referidos, aplicou-se a sintaxe proposta pelo consórcio W3C (2008). Na Figura 5.20 pode observar-se um exemplo dessa sintaxe.

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?x ?name
WHERE { ?x foaf:name ?name }
```

Figura 5.20: Sintaxe de uma consulta em SPARQL (Fonte: W3C (2008)).

Na SPARQL é possível realizar consultas através de quatro formas distintas: i) CONSTRUCT; ii) DESCRIBE ; iii) ASK; iv) SELECT. Apresenta-se na Tabela 5.21 uma breve descrição do tipo de resultado que cada uma delas origina.

Tabela 5.21: Consultas em SPARQL: tipo de resultados obtidos (Adaptado) (W3C, 2008).

CONSTRUCT	Devolve as triples que fazem parte da ontologia.
DESCRIBE	Devolve uma descrição dos recursos encontrados.
ASK	Devolve um valor do tipo booleano (true; false) que indica a veracidade das condições verificadas.
SELECT	Devolve, total ou parcialmente, o valor das variáveis que intervêm na consulta.

Nas consultas que podem ser observadas na secção 6.4 destaca-se ainda a utilização de:

- WHERE (permite a definição de condições);
- ORDER By (utilizado quando se pretende realizar ordenações);
- FILTER (permite a aplicação de filtros);
- LIMIT (limita a visualização dos resultados);
- DISTINCT (na presença de resultados repetidos devolve apenas um resultado);
- GROUP BY (utilizado para criar agrupamento dos dados, associado normalmente a funções).





# 6

## Desenvolvimento das ontologias

### 6.1 Cenário I: transposição do modelo SPARES para ontologia

Nesta abordagem consideraram-se as tabelas fulcrais, no repositório SPARES, para as pesquisas que se preveem vir a ser necessárias. Nesse sentido, a construção da ontologia baseou-se no esquema concetual da base de dados.

#### 6.1.1 Classes

A construção da ontologia foi iniciada com a determinação das classes que lhe servirão de base estrutural. Deste modo, criaram-se três classes: OntRel, OntLoc e OntMin (Figura 6.1). Como se referiu na secção 5.7, estas foram geradas automaticamente pelo Protégé.

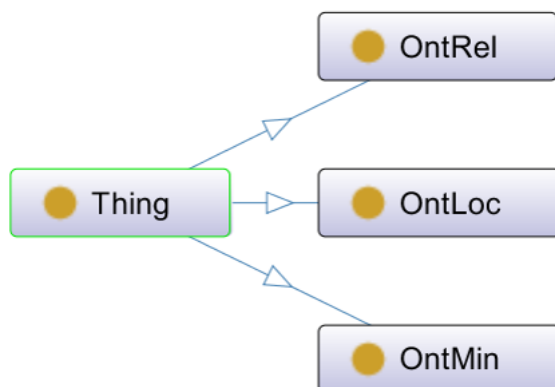


Figura 6.1: Classes desenhadas para o cenário I.

### 6.1.2 Propriedades de tipo de dados

As propriedades de tipo de dados foram criadas automaticamente através da ferramenta *Generate ontology and mappings*. Como se pode observar pela Figura 6.2, as propriedades de tipo de dados não ficam associadas a nenhuma das tabelas importadas. Posteriormente foi feita essa associação (Figura 6.3).

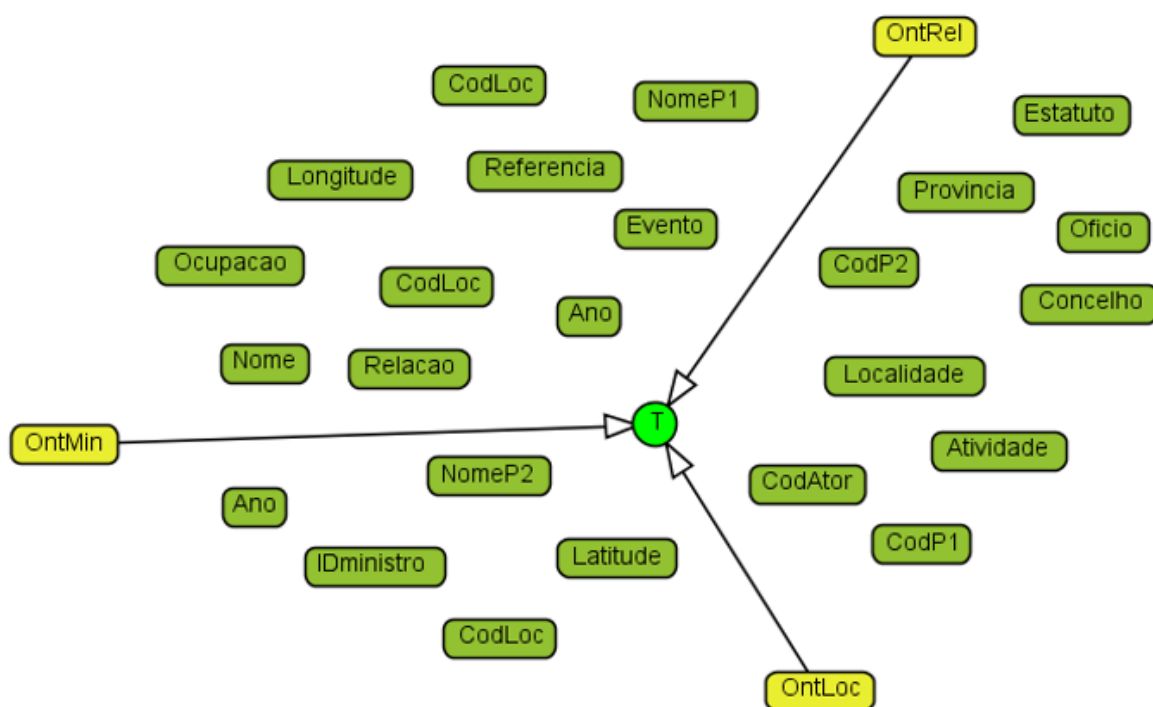


Figura 6.2: Classes e propriedades de tipo de dados obtidas através da ferramenta *Generate ontology and mappings*.

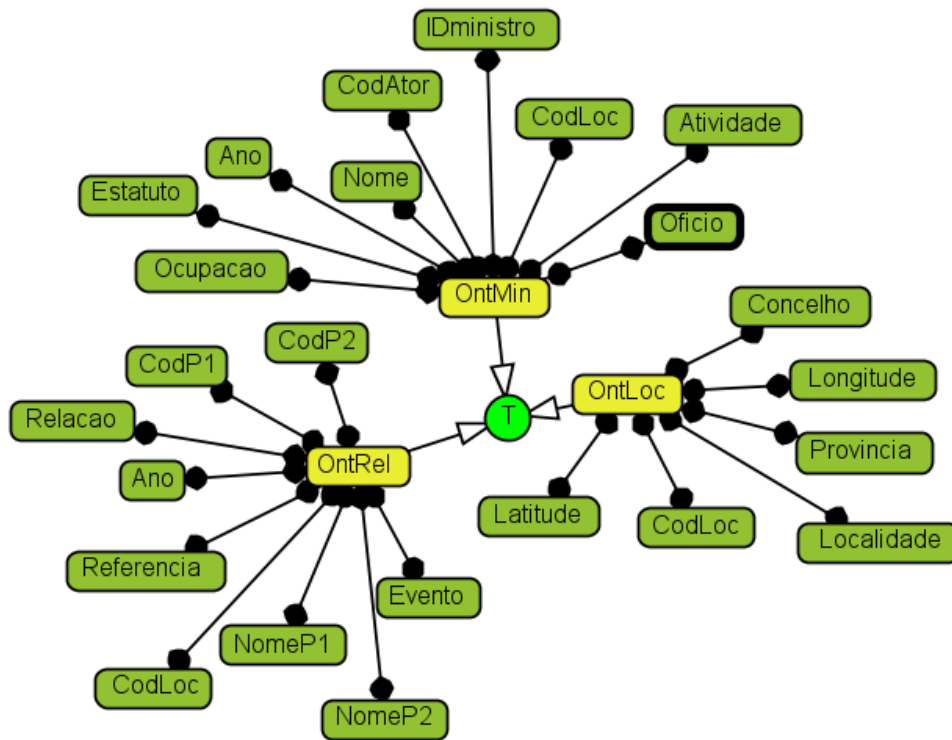


Figura 6.3: Classes e respectivas propriedades de tipo de dados.

Pode observar-se na Tabela 6.1 as propriedades de tipo de dados atrás referidas associadas ao respetivo domínio e contradomínio.

### 6.1.3 Propriedades de objeto

Criaram-se duas propriedades de objeto. Estas permitirão associar a classe Localizacoes às duas outras classes (Tabela 6.2). Foram também criadas as respetivas classes inversas: éLocalizacaoDeMin e éLocalizacaoDeRel.

## 6.2 Cenário II: criação de nova ontologia

Para a construção desta ontologia considerou-se não só o cenário de referência, mas também o conhecimento obtido com a triagem de palavras-chave.

### 6.2.1 Classes

Iniciou-se a construção da ontologia com a criação das quatro classes principais (Figura 6.4).

Apresenta-se nas figuras 6.5, 6.6 e 6.7 as subclasses que foram criadas para três das quatro classes visualizadas na Figura 6.4.

Tabela 6.1: Propriedades de tipo de dados, no cenário I.

Propriedade de tipo de dado	Domínio	Contradomínio
CodLoc	OntLoc	String
Localidade		String
Concelho		String
Provincia		String
Latitude		Double
Longitude		Double
Referencia	OntRel	String
NomeP1		String
CodP1		String
Relacao		String
CodP2		String
NomeP2		String
Ano		Integer
Evento		String
CodLoc		String
IDministro	OntMin	String
Nome		String
Ano		Integer
Oficio		String
Ocupacao		String
Atividade		String
Estatuto		String
CodAtor		String
CodLoc		String

Tabela 6.2: Propriedades de objeto, no cenário I.

Propriedade de objeto	Domínio	Contradomínio
temLocalizacaoMin	OntMin	OntLoc
temLocalizacaoRel	OntRel	OntLoc

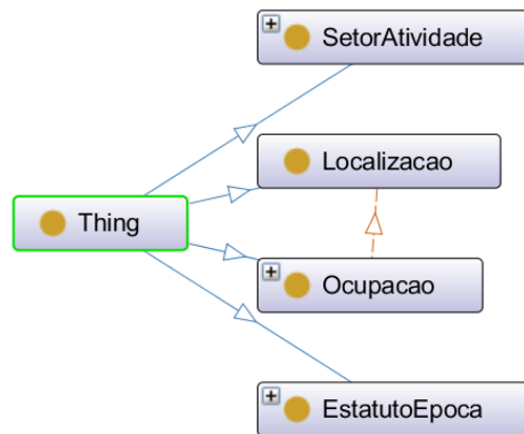


Figura 6.4: Classes principais da ontologia - cenário II.

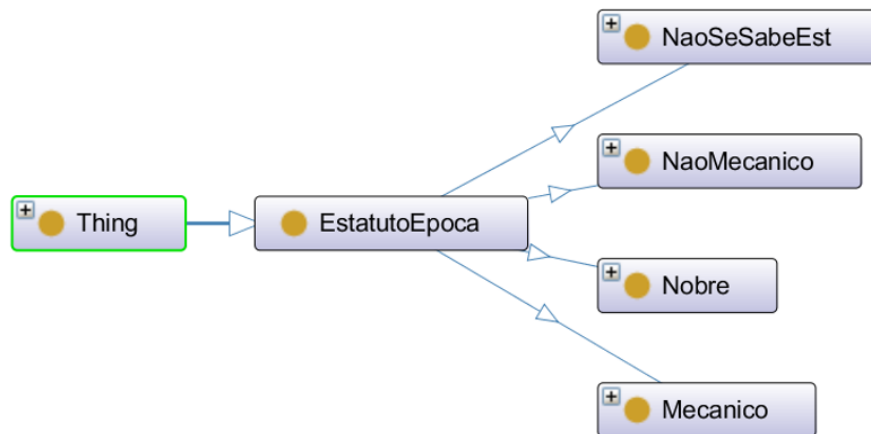


Figura 6.5: Subclasses da classe EstatutoEpoca.

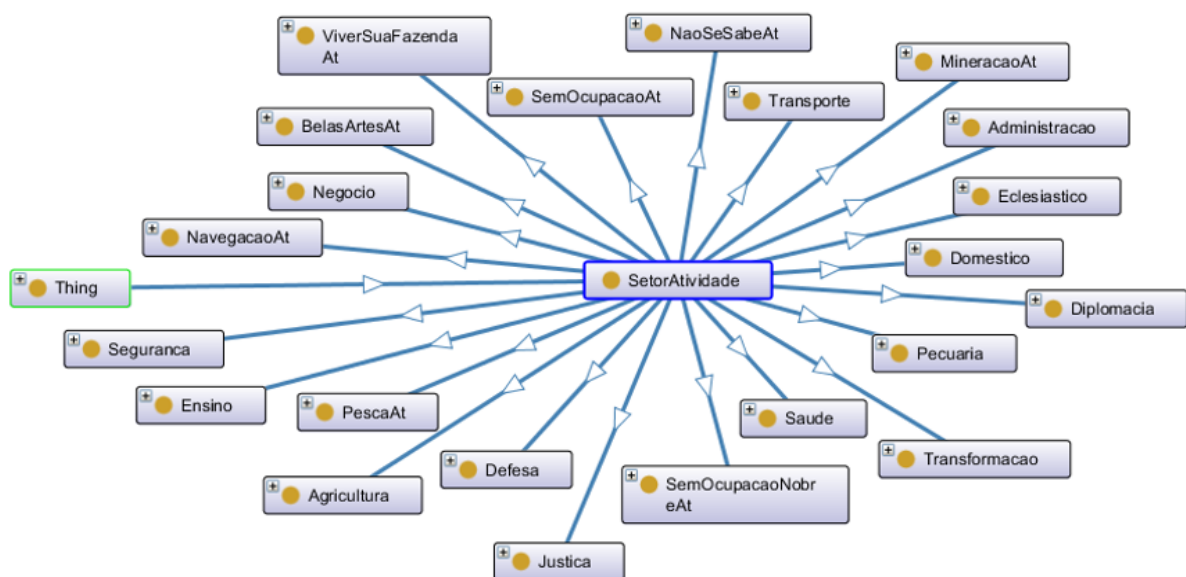


Figura 6.6: Subclasses da classe SetorAtividade.

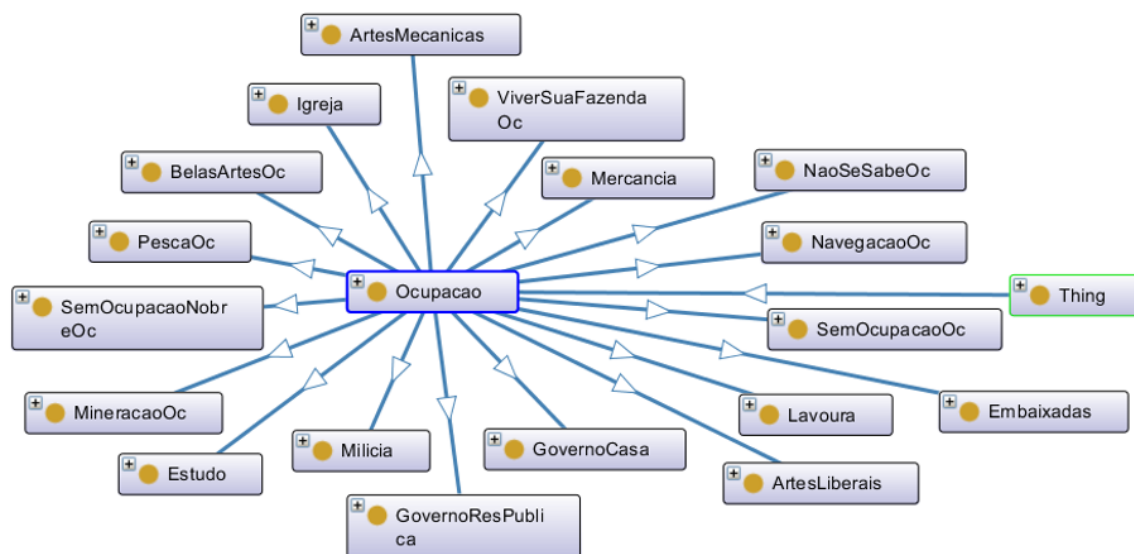


Figura 6.7: Subclasses da classe Ocupacao.

## 6.2.2 Propriedades de tipo de dados

Criaram-se propriedades de tipo de dados com domínio na classe Localizacao e com contradomínio no tipo de dados *double*: temLatitude, temLongitude. Para além destas propriedades, que permitem identificar a localização exata da ocorrência de uma ocupação através das suas coordenadas geográficas, foram ainda criadas as propriedades temLocalidade (nome da localidade), temConcelho (nome do concelho a que pertence a localidade) e temProvincia (nome da província a que o concelho pertence). Estas três propriedades têm como contradomínio o tipo de dados *string*. Com domínio na classe Localizacao criou-se a propriedade temSede, cujo contradomínio é do tipo *boolean* (uma ocupação pode ocorrer numa localidade que é ou não sede de concelho). Associámos à classe Ocupacao propriedades de tipo de dados que permitem uma melhor definição de cada uma das ocupações encontradas na época em estudo. Foram, assim, criadas as propriedades de tipo de dados que se podem consultar na Tabela 6.3.

Tabela 6.3: Propriedades de tipo de dados com domínio na classe Ocupacao.

Propriedade de tipo de dados	Descrição	Contradomínio
temAnoInicial	Ano inicial em que uma dada ocupação foi identificada	integer
temAnoFinal	Ano final em que uma dada ocupação foi identificada	integer
temOficio	Oficio que se encontra associado a uma dada ocupação	string
temLocalidade	Localidade onde determinada ocupação ocorreu	string
temEvento	Evento que corresponde a uma dada ocupação	"Biografico", "Relacional"

## 6.2.3 Propriedades de objeto

Criaram-se as propriedades de objeto necessárias de modo a representar todos os relacionamentos previstos para a ontologia. Na Tabela 6.4 visualizam-se as propriedades cujo domínio são subclasses da classe SetorAtividade e cujo contradomínio são subclasses de Ocupacao.

Tabela 6.4: Propriedades de subclasses de SetorAtividade para subclasses de Ocupação.

Propriedade de objeto	Domínio	Contradomínio
temArtesLiberais	Administracao Ensino Justica Saude	ArtesLiberais
temArtesMecanicas	Ensino Saude Transformacao Transporte	ArtesMecanicas
temBelasArtesOc	BelasArtesAt	BelasArtesOc
temEmbaixadas	Diplomacia	Embaixadas
temEstudo	Ensino	Estudo
temGovernoCasa	Administracao Domestico NaoSeSabeAt	GovernoCasa
temGovernoRespublica	Administracao Defesa Eclesiastico Justica Saude Seguranca Transporte	GovernoResPublica
temIgreja	Eclesiastico	Igreja
temLavoura	Agricultura Pecuaria	Lavoura
temMercancia	Negocio	Mercancia
temMilicia	Defesa Seguranca	Milicia
temMineracaoOc	MineracaoAt	MineracaoOc
temNaoSeSabeOc	Administracao NaoSeSabeAt	NaoSeSabeOc
temNavegacaoOc	NavegacaoAt	NavegacaoOc
temPescaOc	PescaAt	PescaOc
temSemOcupacaoNobreOc	SemOcupacaoNobreAt	SemOcupacaoNobreOc
temSemOcupacaoOc	SemOcupacaoAt	SemOcupacaoOc
temViverSuaFazendaOc	ViverSuaFazendaAt	ViverSuaFazendaOc

A Tabela 6.5 apresenta as propriedades com domínio nas subclasses da classe Ocupacao e com contradomínio nas subclasses de EstatutoEpoca. As propriedades com domínio nas subclasses de SetorAtividade e com contradomínio nas subclasses de EstatutoEpoca podem ser consultadas na Tabela 6.6.

Tabela 6.5: Propriedades de Subclasses de Ocupacao para Subclasses de EstatutoEpoca.

Propriedade de objeto	Domínio	Contradomínio
temMecanico	ArtesMecanicas BelasArtesOc Estudo GovernoCasa GovernoResPublica Lavoura Mercancia Milicia NaoSeSabeOc NavegacaoOc PescaOc	Mecanico
temNaoMecanico	ArtesLiberais ArtesMecanicas BelasArtesOc Estudo GovernoCasa GovernoResPublica Igreja Lavoura Mercancia Milicia MineraçãoOc NaoSeSabeOc NavegacaoOc SemOcupacaoOc SemOcupacaoNobreOc ViverSuaFazendaOc	NaoMecanico
temNaoSeSabeEst	ArtesLiberais ArtesMecanicas BelasArtesOc GovernoCasa GovernoResPublica NaoSeSabeOc	NaoSeSabeEst
temNobre	ArtesLiberais Embaixadas Estudo GovernoCasa GovernoResPublica Igreja Lavoura Milicia NavegacaoOc SemOcupacaoOc SemOcupacaoNobreOc ViverSuaFazendaOc	Nobre



Tabela 6.6: Propriedades de subclasses de SetorAtividade para subclasses de EstatutoEpoca.

Propriedade de objeto	Domínio	Contradomínio
temMecanico	Administracao Agricultura BelasArtesAt Defesa Domestico Ensino Justica NaoSeSabeAt NavegacaoAt Negocio Pecuaria PescaAt Saude Seguranca Transformacao Transporte	Mecanico
temNaoMecanico	Administracao Agricultura BelasArtesAt Defesa Domestico Eclesiastico Ensino Justica MineracaoAt NaoSeSabeAt NavegacaoAt Negocio Saude Seguranca SemOcupacaoAt SemOcupacaoNobreAt Transformacao Transporte ViverSuaFazendaAt	NaoMecanico
temNaoSeSabeEst	Administracao BelasArtesAt Domestico Ensino Saude	NaoSeSabeEst
temNobre	Administracao Agricultura Defesa Diplomacia Domestico Eclesiastico Ensino Justica NavegacaoAt Seguranca SemOcupacaoAt SemOcupacaoNobreAt ViverSuaFazendaAt	Nobre

Para todas as propriedades de objeto referidas nas tabelas 6.4, 6.5 e 6.6 foram definidas propriedades inversas. Estas podem ser consultadas na Figura 6.8. Criaram-se ainda mais duas propriedades de objeto: temLocalizacao e éLocalizacaoDe. Com estas propriedades foi possível relacionar as ocupações com a localização da sua ocorrência. Atribuiu-se a temLocalizacao a classe Ocupacao como domínio e a classe Localizacao como contradomínio. A propriedade éLocalizacaoDe foi definida como inversa de temLocalizacao.

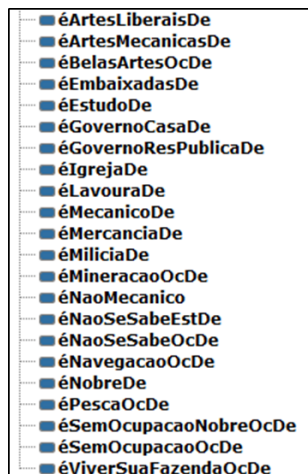


Figura 6.8: Propriedades de objeto inversas.

### 6.3 Inserção de instâncias

Para que os dois cenários possam ser povoados é necessário efetuar a respetiva conexão aos dados e utilizar a ferramenta *Materialize triples*. Exemplifica-se o preenchimento das instâncias com o resultado obtido para a "IDministro=31187", no Cenário I (Figura 6.9), e para "Localizacao-1013", no Cenário II (Figura 6.10).

<ul style="list-style-type: none"> <li>IDministro=16865</li> <li>IDministro=19930</li> <li>IDministro=17634</li> <li>IDministro=18689</li> <li>IDministro=3134</li> <li><b>IDministro=31187</b></li> <li>IDministro=16461</li> <li>IDministro=18244</li> <li>IDministro=17273</li> <li>IDministro=17196</li> <li>IDministro=201</li> <li>IDministro=1914</li> <li>IDministro=10885</li> </ul>	<p>Data property assertions +</p> <ul style="list-style-type: none"> <li>Estatuto "Mecânico"</li> <li>CodAto "20111"</li> <li>Ocupacao "Mercancia"</li> <li>Nome "João Ribeiro de Azevedo"</li> <li>Atividade "Negócio"</li> <li>Ano 1724</li> <li>Oficio "Vive do contrato de comprar e vender cera"</li> <li>IDministro "31187"</li> <li>CodLoc "4228"</li> </ul>
---	---

Figura 6.9: Instância: IDministro=31187.

<ul style="list-style-type: none"> <li>Localizacao-101</li> <li>Localizacao-1010</li> <li>Localizacao-1011</li> <li>Localizacao-1012</li> <li><b>Localizacao-1013</b></li> <li>Localizacao-1014</li> <li>Localizacao-1015</li> <li>Localizacao-1016</li> </ul>	<p>Data property assertions +</p> <ul style="list-style-type: none"> <li>temLatitude "41.366"^^double</li> <li>temSede 1</li> <li>temLocalidade "Felgueiras"</li> <li>temLongitude "-8.1966"^^double</li> <li>temProvincia "Entre Douro e Minho"</li> <li>temConcelho "Felgueiras"</li> </ul>
--	---

Figura 6.10: Instância: Localizacao-1013.

## 6.4 Interrogar a ontologia

Utilizou-se a linguagem SPARQL (W3C, 2008) para dar resposta a questões colocadas no âmbito deste estudo. Enunciam-se na Tabela 6.7 algumas das interrogações que foram realizadas sobre a ontologia, bem como a sua sintaxe em SPARQL.

Tabela 6.7: Consultas realizadas sobre a OntoSPARES.

Consulta 1: Listar todas as triples ordenadas por sujeito.

```
CONSTRUCT { ?s ?p ?o }
WHERE { ?s ?p ?o } ORDER BY ?s
```

Consulta 2: Descrever as localizações de Portalegre.

```
DESCRIBE ?x
WHERE { ?x rdf:type :OntLoc. ?x :Concelho ?Concelho.
FILTER regex(?Concelho, "Portalegre") }
```

Consulta 3: Verificar se o concelho do Entroncamento faz parte das localizações.

```
ASK
WHERE { ?x rdf:type :OntLoc. ?x :Concelho ?Concelho. }
FILTER regex(?Concelho, "Entroncamento") }
```

Consulta 4: Listar as latitudes por código de localidade.

```
SELECT ?CodLoc ?Latitude
WHERE { ?CodLoc rdf:type :OntLoc. ?x :Latitude ?Latitude }
```

Consulta 5: Selecionar as ocupações e respetivos estatutos.

```
SELECT ?Ocupacao ?Estatuto
WHERE { [] a :OntMin ; :Ocupacao ?Ocupacao ; :Estatuto ?Estatuto . }
```

Consulta 6: Selecionar as localidades da provincia da Beira com latitude inferior a 40 graus.

```
SELECT ?Localidade ?Latitude
WHERE { [] a :OntLoc ; :Localidade ?Localidade ; :Latitude ?Latitude ; :Provincia ?Provincia .
FILTER regex(?Provincia, "Beira")
FILTER (?Latitude < 40) }
```

Consulta 7: Selecionar os eventos que começam por Ouvida, ordenados de modo descendente.

```
SELECT ?Eventos
WHERE { [] a :OntRel ; :Evento ?Eventos .
FILTER regex(?Eventos, "^Ouvida") }
ORDER BY DESC (?Eventos)
```

Consulta 8: Selecionar os 2 primeiros eventos que começam por Ouvida, ordenados de modo descendente.

```
SELECT ?Eventos
WHERE { [] a :OntRel ; :Evento ?Eventos .
FILTER regex(?Eventos, "^Ouvida") }
ORDER BY DESC (?Eventos)
LIMIT 2
```

Consulta 9: Selecionar, sem repetir, os ofícios que começam por Vive e cujo estatuto é Mecânico.

```
SELECT DISTINCT ?Oficios
WHERE { [] a :OntMin ; :Oficio ?Oficios ; :Estatuto ?Estatuto .
FILTER regex(?Oficios, "^Vive")
FILTER (?Estatuto = "Mecânico") }
```

Consulta 10: Contabilizar quantos ofícios existem por estatuto.

```
SELECT ?Estatutos (COUNT (?Oficios) as ?Quantidade)
WHERE { [] a :OntMin ; :Oficio ?Oficios ; :Estatuto ?Estatutos . }
GROUP BY (?Estatutos)
```

Pode consultar-se nas figuras 6.11, 6.12, 6.13, 6.14, 6.15, 6.16, 6.17, 6.18, 6.19 e 6.20 os resultados das consultas constantes da Tabela 6.7.

Subject	Predicate	Object
Atividade	rdfs:domain	OntMin
Atividade	rdf:type	owl:DatatypeProperty
Atividade	rdfs:range	xsd:string
Atividade	rdf:type	owl:DatatypeProperty
CodAto	rdf:type	owl:DatatypeProperty
CodAto	rdfs:domain	OntMin

Figura 6.11: Resultado parcial da execução da consulta 1.

Subject	Predicate	Object
CodLoc932	Localidade	"Portalegre"^^<http://www.w3.org/2001/XMLSchema#string>
CodLoc932	rdf:type	owl:NamedIndividual
CodLoc932	rdf:type	OntLoc
CodLoc932	rdf:type	owl:NamedIndividual
CodLoc932	Concelho	"Portalegre"^^<http://www.w3.org/2001/XMLSchema#string>
CodLoc932	rdf:type	owl:NamedIndividual

Figura 6.12: Resultado parcial da execução da consulta 2.

Result
False

Figura 6.13: Resultado da execução da consulta 3.

CodLoc	Latitude
CodLoc498	"38.739"^^<http://www.w3.org/2001/XMLSchema#double>
CodLoc3306	"38.739"^^<http://www.w3.org/2001/XMLSchema#double>
CodLoc5701	"38.739"^^<http://www.w3.org/2001/XMLSchema#double>
CodLoc2510	"39.484"^^<http://www.w3.org/2001/XMLSchema#double>
CodLoc1297	"39.484"^^<http://www.w3.org/2001/XMLSchema#double>
CodLoc142	"39.484"^^<http://www.w3.org/2001/XMLSchema#double>

Figura 6.14: Resultado parcial da execução da consulta 4.

Ocupacao	Estatuto
"Viver da sua fazenda"^^<http://www.w3.org/2001/XMLSchema#string>	"Não mecânico"^^<http://www.w3.org/2001/XMLSchema#string>
"Lavoura"^^<http://www.w3.org/2001/XMLSchema#string>	"Não mecânico"^^<http://www.w3.org/2001/XMLSchema#string>
"Viver da sua fazenda"^^<http://www.w3.org/2001/XMLSchema#string>	"Não mecânico"^^<http://www.w3.org/2001/XMLSchema#string>
"Viver da sua fazenda"^^<http://www.w3.org/2001/XMLSchema#string>	"Nobre"^^<http://www.w3.org/2001/XMLSchema#string>
"Viver da sua fazenda"^^<http://www.w3.org/2001/XMLSchema#string>	"Não mecânico"^^<http://www.w3.org/2001/XMLSchema#string>
"Viver da sua fazenda"^^<http://www.w3.org/2001/XMLSchema#string>	"Não mecânico"^^<http://www.w3.org/2001/XMLSchema#string>

Figura 6.15: Resultado parcial da execução da consulta 5.

Localidade	Latitude
"Ansião"^^<http://www.w3.org/2001/XMLSchema#st	"39.912"^^<http://www.w3.org/2001/XMLSchema#double
"Pedrógão Pequeno"^^<http://www.w3.org/2001/XM	"39.904"^^<http://www.w3.org/2001/XMLSchema#double
"Idanha-a-Nova"^^<http://www.w3.org/2001/XMLSc	"39.92"^^<http://www.w3.org/2001/XMLSchema#double>
"Sertão"^^<http://www.w3.org/2001/XMLSchema#stri	"39.804"^^<http://www.w3.org/2001/XMLSchema#double
"Alvaiázere"^^<http://www.w3.org/2001/XMLSchema	"39.844"^^<http://www.w3.org/2001/XMLSchema#double

Figura 6.16: Resultado da execução da consulta 6.

Eventos
"Ouvida como testemunha no processo-crime SO pelo deputado SO"^^<http://www.w3.org/2001/XMLSchema#string>
"Ouvida como testemunha na habilitação de género pelo vigário"^^<http://www.w3.org/2001/XMLSchema#string>
"Ouvida como testemunha na habilitação SO pelo inquisidor SO"^^<http://www.w3.org/2001/XMLSchema#string>
"Ouvida como testemunha na habilitação SO pelo inquisidor SO"^^<http://www.w3.org/2001/XMLSchema#string>
"Ouvida como testemunha na habilitação SO pelo inquisidor SO"^^<http://www.w3.org/2001/XMLSchema#string>
"Ouvida como testemunha na habilitação SO pelo inquisidor SO"^^<http://www.w3.org/2001/XMLSchema#string>

Figura 6.17: Resultado parcial da execução da consulta 7.

Eventos
"Ouvida como testemunha no processo-crime SO pelo deputado SO"^^<http://www.w3.org/2001/XMLSchema#string>
"Ouvida como testemunha na habilitação de género pelo vigário"^^<http://www.w3.org/2001/XMLSchema#string>

Figura 6.18: Resultado da execução da consulta 8.

Ofícios
"Vive de sua agência"^^<http://www.w3.org/2001/XMLSchema#string>
"Vive da sua fazenda e trata em rendas"^^<http://www.w3.org/2001/XMLSchema#string>
"Vive de alugar panos de seda"^^<http://www.w3.org/2001/XMLSchema#string>
"Vive limpamente do seu ofício"^^<http://www.w3.org/2001/XMLSchema#string>
"Vive do contrato de comprar e vender cera"^^<http://www.w3.org/2001/XMLSchema#string>

Figura 6.19: Resultado da execução da consulta 9.

Estatutos	QuantidadeDeOfícios
"Não mecânico"^^<http://www.w3.org/2001/XMLSchem	"60"^^<http://www.w3.org/2001/XMLSchema#integer>
"Nobre"^^<http://www.w3.org/2001/XMLSchema#string	"6"^^<http://www.w3.org/2001/XMLSchema#integer>
"Mecânico"^^<http://www.w3.org/2001/XMLSchema#stri	"5"^^<http://www.w3.org/2001/XMLSchema#integer>

Figura 6.20: Resultado da execução da consulta 10.



# 7

## Discussão de resultados

As métricas de avaliação referidas na secção 3.2 foram adaptadas e aplicadas aos eventos.

Com efeito, podem visualizar-se na Tabela 7.1 os resultados alcançados para a classificação obtida a partir da determinação de palavras-chave conseguidas de modo manual e na Tabela 7.2 a classificação obtida após remoção de *stopwords* e aplicação de *stemming*. A medida *Precision* foi obtida pela relação entre o nº de eventos relevantes recuperados e o nº de eventos recuperados. A medida *Recall* resultou da relação entre o nº de eventos relevantes recuperados e o nº de eventos relevantes.

Tabela 7.1: Métricas de avaliação aplicadas à classificação com palavras-chave manuais.

<i>Precision</i> (P)	Nº de eventos relevantes recuperados	108	0,923
	Nº de eventos recuperados	117	
<i>Recall</i> (R)	Nº de eventos relevantes recuperados	108	0,679
	Nº de eventos relevantes	159	
<i>F measure</i>	$2(PR)/(P+R)$		0,782

Tabela 7.2: Métricas de avaliação aplicadas à classificação com palavras-chave manuais, após remoção de *stopwords* e aplicação de *stemming*.

<i>Precision</i> (P)	Nº de eventos relevantes recuperados	103	0,904
	Nº de eventos recuperados	114	
<i>Recall</i> (R)	Nº de eventos relevantes recuperados	103	0,648
	Nº de eventos relevantes	159	
<i>F measure</i>	$2(PR)/(P+R)$		0,755

Estes resultados permitem-nos concluir que, para a amostra em estudo, a determinação manual de palavras-chave não permitiu a classificação de todos os eventos. De qualquer modo, considerando que a medida *F measure* se encontra acima de 0,7, este pode ser um caminho que vale a pena explorar.

Pontualmente, e para novas ocorrências, ainda poderá ser necessário recorrer à classificação manual, mas para a grande maioria dos eventos a classificar, a utilização de PLN, através da remoção de *stopwords* e aplicação de *stemming*, revelou-se um importante contributo na classificação da amostra estudada.

Os nossos resultados mostram que as listas de *stopwords* disponíveis para a Língua Portuguesa, tanto as disponíveis na internet como as do PostgreSQL, não são totalmente adequadas ao universo dos dados a classificar. Deste modo, a atualização permanente da lista de *stopwords* é fundamental para a classificação dos dados. Neste sentido é tão importante introduzir na lista novas *stopwords* encontradas como remover desta aquelas que sejam identificadas como palavras relevantes para o contexto em estudo.

Ao comparar os dois cenários ontológicos estudados verificou-se que o primeiro cenário, com uma estrutura mais rápida de construir, embora fosse necessária a associação das propriedades de tipo de dados ao respetivo domínio e contradomínio, obriga a que a base de dados conectada ao Protégé contenha apenas as tabelas que farão parte da ontologia. Por outro lado, também não é possível personalizar as propriedades de tipo de dados que são utilizadas. Todos os campos das tabelas dão origem às referidas propriedades. É, no entanto, mais rápido o povoamento global da ontologia, com a aplicação da ferramenta *Materialize triples*, uma vez que não é necessário a construção de *Mappings*. Em qualquer um dos cenários, o desenvolvimento de consultas em SPARQL permitirá a obtenção de respostas, que originarão mais conhecimento prosopográfico da sociedade portuguesa dos séculos XVI, XVII e XVIII.



# 8

## Conclusões

Neste trabalho apresentámos a metodologia seguida na construção de ontologias de domínio que permitem representar relacionamentos em torno do SO e contribuir para ajudar a classificar indivíduos no que respeita ao seu estatuto social: aristocrata, nobre, intermédio, popular, não se sabe. Para atingir este objetivo foi utilizado o *software* Protégé e construídos dois cenários possíveis. Numa primeira abordagem transpôs-se o modelo conceptual do repositório SPARES para o desenho da ontologia. O segundo cenário teve origem nos conhecimentos adquiridos, não só com os procedimentos de PLN aplicados, mas também com a triagem das palavras-chave.

Em ambos os cenários foi criada a hierarquia de classes, assim como foram definidas as propriedades de objeto necessárias para definir as relações entre as diversas classes. Foram de seguida construídas as propriedades de tipos de dados estabelecidas para as classes Ocupacao e Localizacao. Inserir esta última variável, bem como um referente cronológico, foi essencial para o nosso trabalho ser aceite como pertinente por historiadores. Foi ainda utilizado o *plug-in* Ontop, que permitiu associar os dois cenários ontológicos ao SPARES.

Com a elaboração deste trabalho foi possível:

- classificar com maior rapidez um volume elevado de dados prosopográficos complexos e pouco estruturados;

- descobrir novas relações e erros de inserção/classificação;
- redefinir a lista de *stopwords*;
- disponibilizar um instrumento que permita classificar os dados prosopográficos;
- construir e interrogar ontologias;
- incorporar contributos em projetos, na área das humanidades digitais, do CIDEHUS.

No futuro dever-se-á ponderar a construção de uma ontologia que permita a integração de ambos os cenários.

Como o social é complexo e mais ainda o social com um referente histórico, consideramos que a metodologia apresentada é passível de adaptação a outras realidades descritas em linguagem pouco controlada e com parâmetros de temporalidades.

## 8.1 Desafios futuros

Ao concluir este trabalho enunciamos alguns desafios futuros que pretendemos concretizar:

- Criar e disponibilizar um dicionário normalizado de ocupações/categoria social, interagindo por esta via com o projeto internacional HISCO.
- Conceber uma metodologia que permita, por um lado, quando uma palavra-chave está incorretamente presente na lista seja removida desta e por outro, quando não exista possa ser adicionada à referida lista.
- Desenvolver a lista de *stopwords* otimizando a sua adequação à classificação prosopográfica.
- Ponderar a aplicação de outras métricas de avaliação.
- Construir uma plataforma que permita, por um lado, disponibilizar na WEB os conceitos ontológicos criados e, por outro lado, cartografar as localizações, permitindo a sua visualização espacial.

Relativamente a este último desafio encontramos-nos a desenvolver uma aplicação que possibilite aliar os resultados que obtivemos aos sistemas de informação geográfica. Apresentamos no Apêndice E três exemplos que são um protótipo que pretendemos operacionalizar a curto prazo. As imagens apresentadas nas figuras deste apêndice representam dados que se encontram no intervalo cronológico 1650-1674. Este é um dos oito intervalos (com amplitude de 25 anos) em que se dividiram os dados. Confirma-se deste modo a importância da variável tempo no estudo destes dados prosopográficos.

Na Figura E.1 pode consultar-se a distribuição dos quatro estatutos na província da "Beira". É possível observar que, para esta província, o estatuto Não Mecânico predomina sobre os restantes.

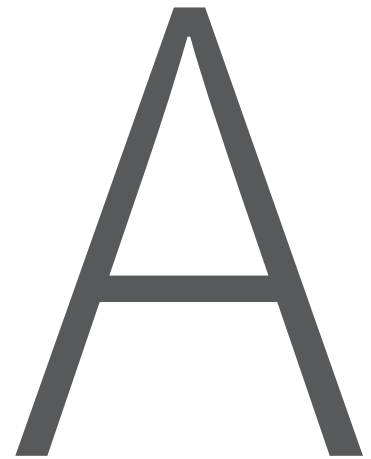
A contabilização dos setores de atividade de uma dada província, e relativamente a uma dada ocupação, é exemplificada na Figura E.2. Assim, na província Estremadura, para a ocupação Artes Mecânicas, sobressai o peso do setor de atividade "Transformação" (80%).

A Figura E.3 destaca a predominância do ofício "Vive de sua fazenda" na província de "Entre Douro e Minho". Evidenciamos nesta projeção uma correlação direta entre este ofício e a densidade populacional desta província.

À semelhança da maior parte das ontologias, este não é um trabalho finalizado, está em permanente construção.

Pretendemos continuar este trabalho utilizando-o como base na descoberta de novas relações em diferentes interpretações dos dados do SO. Assim, a extração de conhecimento proporcionado pelas duas técnicas utilizadas (remoção de *stopwords* e aplicação de *stemming*) e pela ontologia, deverá futuramente facilitar a análise dos relacionamentos existentes entre as ocupações e o estatuto dos diversos membros do SO.





# Publicações consultadas: ontologia e/ou Protégé

(publicadas entre janeiro de 2013 e julho de 2015)

Título	Ano	Área	Protégé *	DOI
A collaborative platform for multilingual ontology development	2014	Multilíngues.	1	
A comprehensive sensor taxonomy and semantic knowledge representation	2013	Sensores.	2	10.1109/ICSensT.2013.6727761
A corpus of OWL DL ontologies	2013	Corpus.	1	
A framework for ocean satellite image classification based on ontologies	2013	Oceanografia.	2	10.1109/JSTARS.2012.2217479
A high-level electrical energy ontology with weighted attributes	2015	Energia.	2	10.1016/j.aei.2015.04.002
A new information model towards context-aware service provisioning in the internet-of-things	2015	E-saúde.	2	10.1109/ICIN.2015.7073827
A novel way to relate ontology classes	2015	Gestão de conferências.	2	10.1155/2015/724196
A query expansion method for retrieving online bim resources based on industry foundation classes	2015	Construção.	2	10.1016/j.autcon.2015.04.006
A research on ontology modeling for multi source heterogeneous audit data	2015	Auditoria.	2	
A review on evolution and versioning of ontology based information systems	2015	Sistemas de informação.	2	10.9790/0661-17233543
A semantic model to query spatial-temporal data	2014	Informação geográfica.	0	10.1007/978-3-642-31833-7_5
A semantic similarity assessment tool for computer science subjects using extended wu & palmer's algorithm and ontology	2015	Educação.	1	10.1007/978-3-662-46578-3_118
A similarity computing method based on a domain ontology for forest pests and diseases	2013	Florestas.	2	
A spatial data model for moving object databases	2014	Informação geográfica.	0	10.5121/ijdms.2013.6101
A study of an agricultural ontology model for an intelligent service in a vertical farm	2013	Agricultura.	2	
A study of various approaches and tools on ontology	2015	Metodologias/conceitos.	1	10.1109/CICT.2015.43
A study on development of cognitive support features in recent ontology visualization tools	2014	Ferramentas de visualização.	2	10.1007/s10462-012-9326-2
A study on the ontology model for preventing livestock disease spread	2013	Zootecnia.	2	
A survey: different approaches to integrate data using ontology and methodologies to improve the quality of data	2014	Qualidade de dados.	0	
A survey on automatic mapping of ontology to relational database schema	2015	Base de dados.	2	
A survey on ontologies for human behavior recognition	2014	Comportamento humano.	0	10.1145/2523819
A survey on user personalization using ontology	2015	Pesquisa na internet.	2	
A survey to service composition methods using aspects classification	2013	Metodologias/conceitos.	0	10.1109/SOSE.2013.55
A synergistic framework for geographic question answering	2013	Informação geográfica.	0	10.1109/ICSC.2013.25
Advances in geographic object-based image analysis with ontologies: a review of main contributions and limitations from a remote sensing perspective	2013	Deteção remota.	1	10.1016/j.isprsjprs.2013.05.003
Aggregating linguistic expert knowledge in type-2 fuzzy ontologies	2015	Indústria.	2	10.1016/j.asoc.2015.03.023
An agent based approach to annotate ideas during creativity challenges in an engineering school of innovation	2014	Educação.	1	
An algorithm and a tool for comparing ontology versions	2013	Medicina.	2	

\* 0 – não referem Protégé; 1 – referenciam Protégé; 2 – utilizam o Protégé

Título	Ano	Área	Protégé *	DOI
An approach for transforming of relational databases to owl ontology	2015	Metodologias/conceitos.	1	10.5121/ijwest.2015.6102
An approach to semantic search engine	2015	Restauração.	2	
An attempt to use ontologies for document image analysis	2014	Análise de documentos.	2	
An extension of protégé for an automatic fuzzy-ontology building using clustering and FCA	2015	Metodologias/conceitos.	2	10.5121/ijcsit.2015.7202
An integrated, ontology-driven approach to constructing observational databases for research	2015	Medicina.	2	10.1016/j.jbi.2015.03.008
An ontological approach for technical plan definition and verification in construction	2015	Construção.	2	10.1016/j.autcon.2015.02.002
An ontology based approach for modeling e-learning in healthcare human resource management	2015	Saúde.	2	
An ontology based information system for the management of institutional repository's collections	2015	Educação.	2	10.1063/1.4907851
An ontology framework for developing platform-independent knowledge-based engineering systems in the aerospace industry	2014	Indústria.	2	10.1080/00207543.2014.919422
An ontology-based approach for inventive problema solving	2014	Física.	2	10.1016/j.engappai.2013.07.005
An ontology-based knowledge representation and implement method for crop cultivation standard	2013	Agricultura.	2	10.1016/j.mcm.2011.11.004
An ontology-based social network analysis prototype	2013	Análise de redes.	2	10.1109/CogSIMA.2013.6523839
An ontology-enabled framework for a geospatial problem-solving environment	2013	Informação geográfica.	1	10.1016/j.compenvurbsys.2012.10.008
A survey: different approaches to integrate data using ontology and methodologies to improve the quality of data	2013	Metodologias/conceitos.	0	
Asking spatial questions to identify GIS functionality	2013	Informação geográfica.	0	10.1109/COMGEO.2013.18
'Aspirational maps' and indigenous communities	2014	Informação geográfica.	0	
Automatic feedback framework for deriving educational ontologies	2014	Educação.	2	10.1109/ICYCS.2008.183
Automatic integration of spatial data in viewing services	2014	Cartografia.	2	
Automatic mapping rules and owl ontology extraction for the obda ontop	2014	Metodologias/conceitos.	2	10.4108/icst.collaboratecom.2014.257493
Automatic transformation of relational database schema into owl ontologies	2015	Metodologias/conceitos.	1	10.17770/etr2015vol3.170
Axioms & templates: distinctions & transformations amongst ontologies, frames, & information models	2013	Metodologias/conceitos.	2	10.1145/2479832.2479840
Bloss: a system for biomedical ontology selection	2014	Medicina.	2	10.1016/j.cmpb.2014.01.020
Bringing agility into linked data development: an industrial use case in logistics domain	2015	Logística.	1	
Building a thematic spatial data infrastructure and situation-aware for global events	2015	Desporto.	2	10.17706/jsw.10.3.304-316
Building an advance domain ontology model of information science (ois)	2014	Ciência da Informação.	2	
Building an ontology in educational domain case study for the university of palestine	2015	Educação.	2	
Building and querying microbial ontology	2013	Microbiologia.	2	10.1016/j.protcy.2013.12.331
Building ontologies for different natural languages	2014	Linguagem natural.	2	10.2298/CSIS130429023A

\* 0 – não referem Protégé; 1 – referenciam Protégé; 2 – utilizam o Protégé



Título	Ano	Área	Protégé *	DOI
Change management in evolving web ontologies	2013	Metodologias/conceitos.	2	10.1016/j.knosys.2012.05.005
Clinical disease registry ontology for blood transfusion	2014	Medicina.	2	10.17758/UR.U1214018
Cognition, ontologies and description logics	2014	Metodologias/conceitos.	2	
Collaborative ontology development for the geosciences	2013	Geociência.	2	10.1111/tgis.12070
Combining 3d models and functions through ontologies to describe man-made products and virtual humans: toward a common framework	2015	Modelação 3D.	2	10.1080/16864360.2014.962429
Comparison some of ontology editors	2013	Metodologias/conceitos.	2	
Concept maps as the first step in na ontology construction method	2013	Metodologias/conceitos.	1	10.1016/j.is.2012.05.010
Conceptology: using concept map for knowledge representation and ontology construction	2013	Tecnologia.	2	10.4304/jnw.8.8.1708-1712
Conceptual and content-based annotation of (multimedia) documents	2014	Documentação.	2	10.1007/s11042-013-1463-3
Conservation of past times: data models for ensuring the future of our heritage	2013	Património cultural.	0	
Constructing conceptual knowledge artefacts: activity patterns in the ontology authoring process	2015	Metodologias/conceitos.	2	10.1145/2702123.2702495
Construction of an ortholog database using the semantic web technology for integrative analysis of genomic data	2015	Biologia.	2	10.1371/journal.pone.0122802
Creation of web 2.0 tools ontology to improve learning	2014	Educação	2	10.1016/j.chb.2014.10.026
Deep analysis for development of rdf, rdfs and owl ontologies with protege	2014	Educação.	2	10.1109/ICRITO.2014.7014747
Developing a process equipment failure knowledge base using ontology approach for process equipment related incidente investigations	2013	Indústria.	2	10.1016/j.jlp.2013.07.013
Developing ontologies and persona to support and enhance requirements engineering activities – a case study	2013	Engenharia de requisitos.	2	10.1016/j.procs.2015.03.060
Developing viso: vaccine information statement ontology for patient education	2015	Medicina.	2	10.1186/s13326-015-0016-2
Development and application of an interaction network ontology for literature mining of vaccine-associated gene-gene interactions	2015	Medicina.	2	10.1186/2041-1480-6-2
Development of an intelligent environmental knowledge system for sustainable agricultural decision support	2014	Agricultura.	0	10.1016/j.envsoft.2013.10.004
Development of ontology in laptop domain for knowledge representation	2015	Metodologias/conceitos.	2	10.1016/j.procs.2015.02.018
Domain ontology based semantic search for efficient information retrieval through automatic query expansion	2013	Metodologias/conceitos.	2	10.1109/ISSP.2013.6526942
Domain ontology driven fuzzy semantic information retrieval	2015	Recuperação de Informação.	2	10.1016/j.procs.2015.02.122
Domain specific answering using semantic approach	2014	Recuperação de Informação.	2	
Dynamic forensics model based on ontology and context information	2013	Judicial.	0	
Effects of guideline-based training on the quality of formal ontologies: a randomized controlled trial	2013	Medicina.	1	10.1371/journal.pone.0061425
Efficient SPARQL-to-SQL with R2RML mappings	2015	Metodologias/conceitos.	2	10.1016/j.websem.2015.03.001
E-learning developing using ontological engineering	2013	E-Learning.	2	

\* 0 – não referem Protégé; 1 – referenciam Protégé; 2 – utilizam o Protégé



Título	Ano	Área	Protégé *	DOI
E-maintenance ontology-based approach for heterogeneous distributed robotic production capabilities	2015	Robótica.	2	10.1080/0951192X.2014.880802
Engineering ontologies for question answering	2014	Metodologias/conceitos.	0	10.3233/AO-140130
Engineering ontology-based access to real-world data sources	2015	Indústria.	2	10.1016/j.websem.2015.03.002
Enhanced methodology for ontology development	2013	Transportes.	0	
Enriching the semantics of variational geometric constraint data with ontology	2015	Desenvolvimento de produto.	2	10.1016/j.cad.2014.12.008
Expert system for medicine diagnosis using software agents	2015	Medicina	2	10.1016/j.eswa.2014.10.026
Exploring ontology-driven modeling approach for multi-agent cooperation in emergency logistics	2014	Serviços de emergência.	2	10.4304/jcp.9.2.285-294
Extracting spatiotemporal and semantic events from documents	2013	Documentação.	0	10.1145/2533888.2533941
Foodwiki: ontology-driven mobile safe food consumption system	2015	Produtos alimentares.	2	10.1155/2015/475410
Formalizing spatiotemporal knowledge in remote sensing applications to improve image interpretation	2013	Deteção remota.	0	10.5311/JOSIS.2013.7.142
Function knowledge retrieval based on functions knowledge ontology	2015	Desenvolvimento de produto.	2	
Generating multiple choice questions from Ontologies: how far can we go?	2014	Educação.	1	10.1007/978-3-319-17966-7_7
Geo-social model: a conceptual framework for real-time geocollaboration	2013	Sociologia.	0	10.1111/j.1467-9671.2012.01352.x
Helping users bootstrap ontologies: an empirical investigation	2015	Metodologias/conceitos.	0	10.1145/2702123.2702433
How ontologies are made: studying the hidden social dynamics behind collaborative ontology engineering projects	2013	Sociologia.	2	10.1016/j.websem.2013.04.001
How to stay on top of your data: Databases, ontologies and more	2015	Metodologias/conceitos.	0	
Impact of an ontology for automatic text classification	2014	Classificação de documentos.	2	
Implementation of linked data in the life sciences at biohackathon 2011	2014	Biologia.	1	10.1186/2041-1480-6-3
Implementation of ontology matching using protégé	2013	Serviços de emergência.	2	
Improved gene ontology annotation for biofilm formation, filamentous growth, and phenotypic switching in candida albicans	2013	Medicina.	0	10.1128 /EC.00238-12
Incorporating ontology-based semantics into conceptual modelling	2015	Medicina.	0	10.1016/j.is.2015.02.003
Indented tree or graph? A usability study of ontology visualization techniques in the context of class mapping evaluation	2013	Metodologias/conceitos.	2	10.1007/978-3-642-41335-3_8
Information modelling in water resources: an australian perspective	2014	Recursos hídricos.	0	10.1007/s00477-013-0754-6
Information retrieval based on semantic navigation in medical corpus	2015	Medicina.	2	
IngeniousTRIZ: an automatic ontology-based system for solving inventive problems	2015	Física.	2	10.1016/j.knosys.2014.11.015
Integrating distributed sources of information for construction cost estimating using semantic web and semantic web service technologies	2015	Tecnologia.	2	10.1016/j.autcon.2015.04.003
Interoperable context sharing in an ontology-enabled collaboration framework	2014	Tecnologia.	2	10.1145/2557977.2558073
Joint: java ontology integrated toolkit	2013	Metodologias/conceitos.	2	10.1016/j.eswa.2013.05.040
Jumping nlp curves: a review of natural language processing research	2014	Linguagem natural.	0	10.1109/MCI.2014.2307227

\* 0 – não referem Protégé; 1 – referenciam Protégé; 2 – utilizam o Protégé

Título	Ano	Área	Protégé *	DOI
Knowledge management systems in the public sector: critical issues	2015	Governamental.	2	10.7763/LNSE.2016.V4.224
Locking for concurrent transactions on ontologies	2013	Metodologias/conceitos.	0	10.1007/978-3-642-38288-8_7
Making the web of data available via web feature services	2014	Informação geográfica.	0	10.1007/978-3-319-03611-3_20
Management of knowledge sources supported by domain ontologies: building and construction case studies	2015	Metodologias/conceitos.	2	10.1002/isaf.1361
Managing water network using GIS	2014	Recursos hídricos.	0	
Mapping evolution of dynamic web ontologies	2015	Metodologias/conceitos.	1	10.1016/j.ins.2014.12.040
Meaningful spatial prediction and aggregation	2014	Estatística.	1	10.1016/j.envsoft.2013.09.00
Methodological guidelines for reusing general ontologies	2013	Metodologias/conceitos.	2	10.1016/j.data.2013.03.006
Mobile cloud-based depression diagnosis using an ontology and a bayesian network	2015	Medicina.	2	10.1016/j.future.2014.05.004
Modeling and representation for earthquake emergency response knowledge: perspective for working with geo-ontology	2014	Serviços de emergência.	2	10.1080/13658816.2013.845893
Modeling of test specifications of raw materials in seafood ontology using semantic web rule language (swrl)	2015	Produtos alimentares.	2	10.1145/2743065.2743066
Modeling smart grid neighborhoods with the enersip ontology	2015	Energia.	2	10.1016/j.compind.2015.01.008
Model-theoretic inseparability and modularity of description logic ontologies	2013	Metodologias/conceitos.	1	10.1016/j.artint.2013.07.004
Multidimensional ontology model to support context- aware systems	2013	Tecnologia.	1	
OBDA: query rewriting or materialization? In practice, both!	2014	Metodologias/conceitos.	0	10.1007/978-3-319-11964-9_34
OBESTDD: ontology based expert system for thyroid disease diagnosis	2015	Medicina.		10.1109/ICNTE.2015.7029914
On the cyberinfrastructure for GIS-enabled historiography	2013	Informação geográfica.	0	10.1080/00045608.2013.792178
Ontological modelling for intelligent e-learning	2014	E-learning.	2	10.1109/ICALT.2014.208
Ontologies for crisis management: a review of state of the art in ontology design and usability	2013	Metodologias/conceitos.	1	
Ontologies for probabilistic situation assessment in the maritime domain	2013	Biologia.	2	10.1109/CogSIMA.2013.6523830
Ontologies: need, usage and attainment of health care system	2013	Saúde.	2	10.1109/ISSP.2013.6526939
Ontology alignment evaluation initiative: six years of experience	2013	Metodologias/conceitos.	0	10.1007/978-3-642-22630-4_6
Ontology and indigeneity: on the political ontology of heterogeneous assemblages	2014	Política.	0	10.1177/1474474012462534
Ontology application and use at the ENCODE DCC	2015	Biologia.	2	10.1093/database/bav010
Ontology based approach for semantic information retrieval system	2015	Metodologias/conceitos.	1	
Ontology based description of an accessible learning object	2013	E-learning.	2	10.1109/ICTA.2013.6815284
Ontology based e-learning approach over traditional e-learning	2014	E-learning.	1	
Ontology based public healthcare system in internet of things (iot)	2015	Medicina.	2	10.1016/j.procs.2015.04.067
Ontology design and individual cognitive peculiarities: a pilot study	2015	Psicologia.	2	10.1016/j.eswa.2015.01.008
Ontology development and query retrieval using protégé tool	2013	Metodologias/conceitos.	2	10.5815/ijisa.2013.09.0
Ontology engineering in the ERA of linked data	2015	Metodologias/conceitos.	1	10.1002/bult.2015.1720410407
Ontology evolution and challenges	2013	Metodologias/conceitos.	2	
Ontology for feature based selection of web development tools	2013	Tecnologia.	2	10.1109/ICDIM.2013.6693980
Ontology in association rules	2013	Data mining.	1	10.1186/2193-1801-2-452

\* 0 – não referem Protégé; 1 – referenciam Protégé; 2 – utilizam o Protégé

Título	Ano	Área	Protégé *	DOI
Ontology integration for advanced manufacturing collaboration in cloud platforms	2015	Tecnologia.	2	10.1109/INM.2015.7140329
Ontology mapping using description logic and bridging axioms	2015	Empresarial.	2	10.1016/j.compind.2012.09.004
Ontology matching: state of the art and future challenges	2013	Metodologias/conceitos.	1	10.1109/TKDE.2011.253
Ontology modelling methodology for temporal and interdependent applications	2014	Metodologias/conceitos.	0	10.1109/UKSim.2014.61
Ontology partitioning: clustering based approach	2015	Metodologias/conceitos.	0	10.5815/ijitcs.2015.06.01
Ontology patterns: clarifying concepts and terminology	2013	Metodologias/conceitos.	0	
Ontology-assisted systems engineering process with focus in the requirements engineering process	2015	Engenharia de processos.	0	10.1007/978-3-319-11617-4_11
Ontology-aware partitioning for knowledge graph identification	2013	Metodologias/conceitos.	0	10.1145/2509558.2509562
Ontology-based approach for in vivo human connectomics: the medial brodmann area 6 case study	2015	Medicina.		10.3389/fninf.2015.00009
Ontology-based approach for structural design considering low embodied energy and carbon	2015	Energia.	2	10.1016/j.enbuild.2015.04.051
Ontology-based classification of building types detected from airborne laser scanning data	2014	Construção.	2	10.3390/rs6021347
Ontology-based data access: ontop of databases	2013	Metodologias/conceitos.	2	10.1007/978-3-642-41335-3_35
Ontology-based disassembly information system for enhancing disassembly planning and design	2015	Metodologias/conceitos.	2	10.1007/s00170-014-6704-8
Ontology-based economic models for bioenergy and biofuel projects	2015	Energia.	2	10.1109/ICOSC.2015.7050839
Ontology-based executable design decision template representation and reuse	2015	Metodologias/conceitos.	2	
Ontology-based feature modeling: an empirical study in changing scenarios	2015	Metodologias/conceitos.	2	10.1016/j.eswa.2015.02.020
Ontology-based methodology for managing heterogeneous wireless sensor networks	2013	Tecnologia.	2	10.1155/2013/610684
Ontology-based neural network for patent knowledge management in design collaboration	2013	Tecnologia.	2	10.1080/00207543.2012.701775
Ontology-based privacy preserving digital forensics framework	2015	Judicial.	2	10.14257/ijisia.2015.9.4.07
Ontology-based quantitative similarity metric for event matching in publish/subscribe system	2015	Transportes.	2	10.1016/j.neucom.2014.11.017
Ontology-based semantic interpretation of cylindricity specification in the next-generation GPS	2015	Indústria.	2	10.1016/j.procir.2015.04.054
Ontology-based semantic modeling of construction safety knowledge: towards automated safety planning for job hazard analysis (JHA)	2015	Segurança no trabalho.	2	10.1016/j.autcon.2015.02.005
Ontology-based semantic retrieval for mechanical design knowledge	2015	Desenho de projetos.	2	10.1080/0951192X.2013.874593
Ontology-based text classification into dynamically defined topics	2014	Metodologias/conceitos.	0	10.1109/ICSC.2014.5
Ontology-based topological representation of remote sensing images	2014	Deteção remota.	0	10.1080/01431161.2013.858847
Ontology-driven slope modeling for disaster management service	2015	Florestas.	2	10.1007/s10586-015-0424-1
Ontop at work	2013	Metodologias/conceitos.	0	

\* 0 – não referem Protégé; 1 – referenciam Protégé; 2 – utilizam o Protégé



Título	Ano	Área	Protégé *	DOI
ONTORAT: automatic generation of new ontology Terms, annotations, and axioms based on ontology Design patterns	2015	Medicina.	2	10.1186/2041-1480-6-4
ONTOSOC: sociocultural knowledge ontology	2015	Sociologia.	2	10.5121/ijwest.2015.6201
OPPL-galaxy, a galaxy tool for enhancing ontology exploitation as part of bioinformatics workflows	2013	Bioinformática.	2	10.1186/2041-1480-4-2
Optimising resolution-based rewriting algorithms for owl ontologies	2015	Metodologias/conceitos.	0	10.1016/j.websem.2015.02.001
OPTIQUE – zooming in on big data access	2014	Indústria.	0	10.1109/MC.2015.82
Organization of the concepts of the platonic dialogue parmenides into a software ontology	2015	Metodologias/conceitos.	2	10.1063/1.4907832
OWL 2 based validation and modeling of logistic domain ontology	2014	Logística.	2	10.14257/ijaias.2014.2.2.01
Prospects for a space-time GIS	2013	Informação geográfica.	0	10.1080/00045608.2013.792175
Public transport ontology for passenger information retrieval	2014	Transportes.	2	
Qualitative GIS and the visualization of narrative activity space data	2013	Informação geográfica.	0	10.1080/13658816.2012.678362
RDF and OWL ontology building of web applications	2013	Metodologias/conceitos.	2	
Research on construction method of agricultural domain ontology	2015	Agricultura.	2	
Research on key techniques for ontology construction system	2015	Metodologias/conceitos.	0	
Research on multi-level association rules based on geosciences data	2013	Geociência.	0	10.4304/jsw.8.12.3269-3276
Research on representation and retrieval for equipment fault information based on domain ontology	2015	Transportes.	2	10.2991/icisim.15.2015.69
Revisiting ontologies: a necessary clarification	2013	Metodologias/conceitos.	0	10.1002/asi.22861
Rules and ontology based data access	2014	Metodologias/conceitos.	2	10.1007/978-3-319-11113-1_11
Semantic information retrieval based on domain ontology	2015	Metodologias/conceitos.	2	
Semantic mediator querying	2014	Metodologias/conceitos.	0	10.1145/2628194.2628218
Semantic queries in distributed relational database using global ontology construction	2015	Web semântica.	2	
Semantic solutions for the digital libraries based on semantic web technologies	2014	Documentação.	1	
Semantic web and service computation in giscience applications: a perspective and prospective	2013	Web semântica.	2	10.1080/10106049.2013.776644
Semantic webquery on e-governance data and designing ontology for agriculture domain	2013	Agricultura.	2	10.5121/ijwest.2013.4307
Semantic web technology and ontology designing for e-learning environments	2015	E-Learning.	2	
Semantically enhanced decision support for learning management systems	2015	Educação.	2	10.1109/ICOSC.2015.7050823
Semantic-based approach for route determination and ontology updating	2013	Transportes.	2	10.1016/j.engappai.2012.09.012
Semantics-based services for a low carbon society: an application on emissions trading system data and scenarios management	2015	Ambiente.	2	10.1016/j.envsoft.2014.11.007
Semi-automatic construction of a domain ontology for wind energy using wikipedia articles	2014	Energia.	2	10.1016/j.renene.2013.08.002
Semi-automatic ontology merging of domain specific ontologies	2015	Metodologias/conceitos.	2	
SemRPer - a rule based personalization system for semantic web	2015	Metodologias/conceitos.	2	

\* 0 – não referem Protégé; 1 – referenciam Protégé; 2 – utilizam o Protégé

Título	Ano	Área	Protégé *	DOI
Sentiments analysis at conceptual level making use of the narrative knowledge representation language	2014	Metodologias/conceitos.	1	10.1016/j.neunet.2014.05.010
Sextant: browsing and mapping the ocean of linked geospatial data	2013	Informação geográfica.	0	10.1007/978-3-642-41242-4_26
Social analytics: learning fuzzy product ontologies for aspect-oriented sentiment analysis	2014	Metodologias/conceitos.	0	10.1016/j.dss.2014.05.005
Social learning network analysis model to identify learning patterns using ontology clustering techniques and meaningful learning	2013	E-Learning.	2	10.1016/j.compedu.2012.11.011
Social networking ontology engineering walkthrough: practical approach for non-expert user learning	2015	Redes sociais.	2	10.1007/s11036-014-0559-y
Source code plagiarism detection method using protégé built ontologies	2013	Metodologias/conceitos.	2	10.12948/issn14531305/17.3.2013.07
Space-time representation and analytics	2014	Informação geográfica.	0	10.1080/19475683.2013.862301
Spatial knowledge ontologies	2014	Metodologias/conceitos.	0	10.5829/idosi.wasj.2014.31.02.14310
Spatio-temporal model for role assignment in wireless sensor networks	2013	Tecnologia.	2	
Storing OWL ontologies in object-oriented databases	2015	Metodologias/conceitos.	2	10.1016/j.knosys.2014.12.020
Supporting parent's at NICU using automatic summary generator system - NICU ontology	2015	Medicina.	2	10.14257/ijunesst.2015.8.3.04
Temporal logic and operation relations based knowledge representation for land cover change web services	2013	Metodologias/conceitos.	2	10.1016/j.isprsjprs.2013.02.005
Temporal three-dimensional ontology for geographical information science (GIS)—a review	2013	Modelação 3D.	0	10.4236/ijgis.2013.53030
The basic formal ontology as a reference framework for modeling the evolution of administrative units	2013	Administração pública.	2	10.1111/j.1467-9671.2012.01356.x
The data mining optimization ontology	2015	Data mining.	2	10.1016/j.websem.2015.01.001
The drosophila anatomy ontology	2013	Anatomia.	2	10.1186/2041-1480-4-32
The neurological disease ontology	2013	Medicina.	2	10.1186/2041-1480-4-42
The NPD benchmark: reality check for OBDA systems	2015	Indústria.	1	10.5441/002/edbt.2015.62
The ontop framework for ontology based data access	2014	Metodologias/conceitos.	2	10.1007/978-3-662-45495-4_6
The price of query rewriting in ontology-based data access	2014	Metodologias/conceitos.	0	10.1016/j.artint.2014.04.004
The protégé VOWL plugin: ontology visualization for everyone	2014	Ferramentas de visualização.	2	
The research on reasoning of tourism information and knowledge based on semantic web	2015	Turismo.	2	
Thoughts on the complex relation between linked data, semantic annotations, and ontologies	2013	Web semântica.	0	10.1145/2513204.2513218
Towards a unified notation for owl ontologies: insights from a comparative user study	2013	Metodologias/conceitos.	2	10.1145/2506182.2506192
Towards an intelligent livestock farm management using OWL- based ontology model	2015	Zootecnia.	2	10.14257/ijsh.2015.9.4.25
Transportation ontology definition and application for the content personalization of user interfaces	2013	Transportes.	2	10.1016/j.eswa.2012.12.028
Type-2 fuzzy ontology-based semantic knowledge for collision avoidance of autonomous underwater vehicles	2015	Transportes.	2	10.1016/j.ins.2014.10.013
Unified data model for wireless sensor network	2015	Sensores.	2	10.1109/JSEN.2015.2393951

\* 0 – não referem Protégé; 1 – referenciam Protégé; 2 – utilizam o Protégé

<b>Título</b>	<b>Ano</b>	<b>Área</b>	<b>Protégé *</b>	<b>DOI</b>
Univpeopleprogram ontology: a OWL based structural definition for semantic web	2013	Educação.	2	10.1109/CICT.2013.6558198
Using ontologies to capture the semantics of a (business) process for digital preservation	2015	Documentação.	2	10.1007/s00799-015-0141-7
Using ontologies to perform threat analysis and develop defensive strategies for mobile security	2015	Segurança.	2	10.1007/s10799-014-0213-1
Using ontologies understanding the user experience	2014	Metodologias/conceitos.	2	10.1007/978-3-319-13704-9_43
Using SWRL and Protégé 4.1 to optimize and reason with policies of the cooperation between goal and process	2014	Metodologias/conceitos.	2	10.4304/jsw.9.2.409-416
Virtual leodium: from an historical 3d city scale model to an archaeological information system	2013	Arqueologia.	0	10.5194/isprsannals-II-5-W1-241-2013
VOWL 2: user-oriented visualization of ontologies	2014	Ferramentas de visualização.	2	10.1007/978-3-319-13704-9_21
Web-based ontology editor enhanced by property value extraction	2013	Metodologias/conceitos.	2	10.5121/ijwest.2013.4301
What are ontologies good for?	2013	Metodologias/conceitos.	2	10.1007/978-3-642-34997-3_9
What one can learn from extracting owl ontologies from a netlogo model that was not designed for such an exercise	2015	Judicial.	2	

\* 0 – não referem Protégé; 1 – referenciam Protégé; 2 – utilizam o Protégé

# B

Palavras-chave obtidas  
manualmente



Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
abade	Igreja	Eclesiástico		x			
abadessa	Igreja	Eclesiástico		x			
abegão	Lavoura	Pecuária	x				
abridor; armas	Artes mecânicas	Transformação	x				
açafata	Governo da casa	Doméstico	x				
acamaxeiro	Governo da casa	Transformação	x				
adjutório	Governo da casa	Transformação	x				x
administra; morgadio	Viver da sua fazenda	Viver da sua fazenda		x			
administra; morgado	Viver da sua fazenda	Viver da sua fazenda		x			
administra; negócios	Mercancia	Negócio		x			
administrador; capela	Viver da sua fazenda	Viver da sua fazenda		x			
administrador; convento	Governo da casa	Administração		x			
administrador; correio	Governo da res publica	Administração		x			
administrador; escravos	Mercancia	Negócio	x				
administrador; fazendas	Lavoura	Agricultura		x			
administrador; hospital	Governo da res publica	Administração		x			
administrador; paço	Governo da res publica	Administração		x			
administrador; pólvora	Governo da res publica	Administração		x			
administrador; procurador	Governo da res publica	Administração		x			
administrador; sabão	Mercancia	Negócio		x			
administrador; tabaco	Mercancia	Negócio		x			
adobeiro	Artes mecânicas	Transformação	x				
advoga	Artes liberais	Justiça		x			
advogado	Artes liberais	Justiça		x			
advogando	Artes liberais	Justiça		x			
advogar	Artes liberais	Justiça		x			
agente; cabido	Artes liberais	Administração		x			
agente; causas	Artes liberais	Administração		x			
agente; cousas	Governo da res publica	Administração		x			
agente; duque	Governo da casa	Administração		x			
agente; fisco	Governo da res publica	Administração		x			
agente; mar; guerra	Milícia	Defesa	x				
agente; negócio	Mercancia	Negócio		x			
agente; reservatário	Artes liberais	Administração		x			
agente; universidade	Artes liberais	Administração		x			
agostinho	Igreja	Eclesiástico		x			
aguadeiros	Artes mecânicas	Negócio	x				
aguardenteiro	Artes mecânicas	Transformação	x				
agulheiro	Artes mecânicas	Transformação	x				
ajudante	Milícia	Defesa	x				
ajudante; secretário	Governo da res publica	Administração		x			
ajudante; secreto	Governo da res publica	Administração		x			
albardeiro	Artes mecânicas	Transformação	x				
alcaide	Governo da res publica	Defesa		x			
alcaide; cárceres	Governo da res publica	Administração		x			
alfaiate	Artes mecânicas	Transformação	x				
alferes	Milícia	Defesa		x			
alferes; ajudante	Milícia	Defesa	x				
alferes; chanceler	Governo da res publica	Administração		x			
alferes; reformado	Milícia	Defesa		x			
alferes-mor	Milícia	Defesa		x			
algebrista-mestre	Artes mecânicas	Saúde	x				
algibebe	Artes mecânicas	Transformação	x				
almirante	Navegação	Navegação		x			
almocreve	Artes mecânicas	Transporte	x				
almoedeiro	Artes mecânicas	Transformação	x				
almotacé	Governo da res publica	Administração		x			
almoxarife	Governo da res publica	Administração		x			
alquilador	Mercancia	Negócio	x				

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe



Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
altareiro	Artes mecânicas	Transformação	x				
alugador; baetas	Mercancia	Negócio	x				
alugar; panos	Mercancia	Negócio	x				
alvanel	Artes mecânicas	Transformação	x				
alvitador	Artes mecânicas	Transformação	x				
ama	Governo da casa	Doméstico	x				
amanha; vinhas	Lavoura	Agricultura	x				
amassador	Artes mecânicas	Transformação	x				
andador	Artes mecânicas	Transporte	x				
anzoleiro	Artes mecânicas	Transformação	x				
apontador	Artes mecânicas	Transformação	x				
aprovado; medicina	Artes liberais	Saúde		x			
aprovado; servir; fisco	Governo da res publica	Administração		x			
arcebispo	Igreja	Eclesiástico		x			
arcediogo	Igreja	Eclesiástico		x			
arcipreste	Igreja	Eclesiástico		x			
areeiro	Artes mecânicas	Transformação	x				
armador	Artes mecânicas	Transformação	x				
armador; igrejas	Artes mecânicas	Transformação	x				
armador; sardinha	Artes mecânicas	Pesca	x				
armas-oficial	Não se sabe	Não se sabe	x				
armeiro	Artes mecânicas	Transformação	x				
arquitecto	Belas artes	Belas artes	x				
arrenda	Mercancia	Negócio	x				
arrieiro	Artes mecânicas	Transformação	x				
artilheiro	Milícia	Defesa	x				
assentista	Mercancia	Negócio		x			
assessor; arquiduke	Governo da res publica	Administração		x			
assistente	Governo da casa	Administração	x				
assistente; auditórios	Artes liberais	Justiça		x			
assistente; deputado	Governo da res publica	Administração		x			
atafoneiro	Artes mecânicas	Transformação	x				
auditor	Governo da res publica	Administração		x			
avaliador	Artes mecânicas	Transformação	x				
azulador	Artes mecânicas	Transformação	x				
banqueiro	Mercancia	Negócio		x			
barbeiro	Artes mecânicas	Transformação	x				
barqueiro	Artes mecânicas	Transporte	x				
batedor	Artes mecânicas	Transformação	x				
bate-folha	Artes mecânicas	Transformação	x				
bedel	Governo da res publica	Administração	x				
beneficiado	Igreja	Eclesiástico		x			
bergante	Não se sabe	Não se sabe				x	
bispo	Igreja	Eclesiástico			x		
boeiro	Lavoura	Pecuária	x				
borracheiro	Artes mecânicas	Transformação	x				
boticário	Artes mecânicas	Saúde	x				
burocrata	Governo da res publica	Administração		x			
cabeiro	Artes mecânicas	Transformação	x				
cabeleireiro	Artes mecânicas	Transformação	x				
cabo	Milícia	Defesa	x				
cabouqueiro	Artes mecânicas	Transformação	x				
cabreiro	Lavoura	Pecuária	x				
caçador	Lavoura	Pecuária	x				
cadete	Milícia	Defesa				x	
caixeiro	Mercancia	Negócio	x				
calçador	Artes mecânicas	Transformação	x				
calceteiro	Artes mecânicas	Transformação	x				
caldeireiro	Artes mecânicas	Transformação	x				

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe

Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
camarista	Governo da casa	Doméstico		x			
caminheiro	Artes mecânicas	Transporte	x				
campainheiro	Igreja	Administração	x				
canastreiro	Artes mecânicas	Transformação	x				
canteiro	Artes mecânicas	Transformação	x				
cantor; sé	Igreja	Eclesiástico		x			
capelão	Igreja	Eclesiástico		x			
capinha; colegiada; guimarões	Igreja	Eclesiástico		x			
capitão	Milícia	Defesa		x			
capucho	Milícia	Defesa		x			
caramelo	Artes mecânicas	Transformação	x				
carcereiro	Governo da res publica	Justiça	x				
cardador	Artes mecânicas	Transformação	x				
cardeal	Igreja	Eclesiástico			x		
carmelita	Igreja	Eclesiástico		x			
carpinteiro	Artes mecânicas	Transformação	x				
carretas	Artes mecânicas	Transporte	x				
carreteiro	Artes mecânicas	Transporte	x				
carteiro	Artes mecânicas	Transporte	x				
cartório	Governo da res publica	Administração	x				
caseira	Lavoura	Agricultura	x				
catedrático	Estudo	Ensino		x			
cavaqueiro	Artes mecânicas	Transformação	x				
ceareiro	Lavoura	Agricultura	x				
cergueiro	Artes mecânicas	Transformação	x				
cerieira	Artes mecânicas	Transformação	x				
cerieiro	Artes mecânicas	Transformação	x				
cerigueiro	Artes mecânicas	Transformação	x				
chanceler	Governo da res publica	Administração		x			
chante	Igreja	Eclesiástico		x			
chapeleiro	Artes mecânicas	Transformação	x				
chapinheiro	Artes mecânicas	Transformação	x				
chameleiro	Artes mecânicas	Belas artes	x				
cinteiro	Artes mecânicas	Transformação	x				
cirurgião		Saúde					x
cirurgião; aprovado	Artes mecânicas	Saúde	x				
cirurgião-mor	Artes liberais	Saúde		x			
clérigo	Igreja	Eclesiástico		x			
coadjutor	Igreja	Eclesiástico		x			
coadjutor	Igreja	Eclesiástico		x			
cobra; dinheiro	Governo da casa	Administração	x				
cobra; receita	Governo da casa	Administração	x				
cobra; rendimento	Governo da casa	Administração	x				
cobrança	Governo da casa	Administração	x				
cobrava; tercenária	Governo da casa	Administração	x				
colegial	Estudo	Ensino		x			
colmeiro	Artes mecânicas	Transformação	x				
comissário; ad hoc	Igreja	Administração		x			
comissário; bula	Igreja	Administração		x			
comissário; fazendas	Mercancia	Negócio	x				
comissário; mc	Governo da res publica	Administração		x			
comissário; ordens militares	Governo da res publica	Administração		x			
comissário; so	Igreja	Administração		x			
comissário; trigos	Mercancia	Negócio	x				
compra; para; vender	Mercancia	Negócio	x				
comprador	Mercancia	Negócio	x				
comprou; fazendas	Viver da sua fazenda	Viver da sua fazenda		x			

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe

Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
condestável; armadas	Navegação	Navegação		x			
condestável; fortaleza	Milícia	Defesa	x				
condestável; forte	Milícia	Defesa	x				
condutário	Estudo	Ensino		x			
cónego	Igreja	Eclesiástico		x			
confeiteira	Artes mecânicas	Transformação	x				
confeiteiro	Artes mecânicas	Transformação	x				
confessor	Igreja	Eclesiástico		x			
conselheiro	Governo da res publica	Administração		x			
conselheiro; deputado; artes; universidade	Estudo	Ensino		x			
conselho; geral	Governo da res publica	Administração		x			
conserta; canos	Artes mecânicas	Transformação	x				
conservador	Governo da res publica	Administração		x			
conserveira	Artes mecânicas	Transformação	x				
conserveiro	Artes mecânicas	Transformação	x				
consul	Governo da res publica	Administração	x				
consultor	Governo da res publica	Administração		x			
contador	Governo da res publica	Administração	x				
contador-mor	Governo da res publica	Administração		x			
conteiro	Artes mecânicas	Transformação	x				
contínuo	Governo da res publica	Administração	x				
contramestre	Navegação	Navegação	x				
contratador	Mercancia	Negócio		x			
contratante	Mercancia	Negócio	x				
contratava	Mercancia	Negócio	x				
cordoeiro	Artes mecânicas	Transformação	x				
coreiro	Igreja	Eclesiástico		x			
corista	Igreja	Eclesiástico		x			
coronel	Milícia	Defesa		x			
coronheiro	Artes mecânicas	Transformação	x				
correeiro	Artes mecânicas	Transformação	x				
corregedor	Governo da res publica	Administração		x			
correio	Artes mecânicas	Transporte	x				
correio-mor	Governo da res publica	Administração		x			
corretor	Mercancia	Negócio	x				
corretor-mor	Mercancia	Negócio		x			
cortador	Artes mecânicas	Negócio	x				
costureira	Artes mecânicas	Transformação	x				
couteiro	Governo da res publica	Segurança	x				
cozinheiro	Artes mecânicas	Transformação	x				
cravador	Artes mecânicas	Transformação	x				
criada	Governo da casa	Doméstico	x				
criado	Governo da casa	Doméstico	x				
criado; grande	Governo da casa	Doméstico	x				
criado; grave	Governo da casa	Doméstico	x				
criador; cavalos	Lavoura	Pecuária	x				
cultiva; fazendas	Lavoura	Agricultura	x				
cura	Igreja	Eclesiástico		x			
curador	Governo da casa	Justiça		x			
curador; órfãos	Governo da casa	Justiça		x			
curraleiro	Lavoura	Pecuária	x				
cursante	Estudo	Ensino		x			
cursor	Governo da res publica	Transporte	x				
cursou	Estudo	Ensino		x			
curtia; couro	Artes mecânicas	Transformação	x				
curtidor	Artes mecânicas	Transformação	x				
custódio	Igreja	Eclesiástico		x			
cutileiro	Artes mecânicas	Transformação	x				

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe

Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
da; congregação	Igreja	Eclesiástico		x			
da; custódia	Igreja	Eclesiástico		x			
deão	Igreja	Eclesiástico		x			
debulhador	Artes mecânicas	Transformação	x				
definidor	Igreja	Eclesiástico		x			
depositário	Governo da casa	Segurança					x
deputado	Governo da res publica	Administração		x			
desconhecemos; cargo	Não se sabe	Não se sabe				x	
desembargador	Governo da res publica	Justiça		x			
desembargo; real	Governo da res publica	Administração		x			
despenseiro	Mercancia	Negócio	x				
diácono	Igreja	Eclesiástico		x			
distribuidor	Governo da res publica	Administração		x			
dizimeiro	Mercancia	Negócio	x				
do; conselho	Governo da res publica	Administração		x			
dominicano	Igreja	Eclesiástico		x			
donato	Igreja	Eclesiástico	x				
dono; lagar	Lavoura	Transformação	x				
dourador	Artes mecânicas	Transformação	x				
eclesiástico	Igreja	Eclesiástico		x			
ecónomo	Igreja	Eclesiástico	x				
ecónomo; curado	Igreja	Eclesiástico		x			
educanda	Igreja	Ensino		x			
efectua; baptismo	Igreja	Eclesiástico		x			
efectua; casamento	Igreja	Eclesiástico		x			
efectua; óbito	Igreja	Eclesiástico		x			
efetua; curativos	Artes liberais	Saúde		x			
embaixador	Embaixadas	Diplomacia		x			
encadeador; contas	Artes mecânicas	Transformação	x				
enchambrador	Artes mecânicas	Transformação	x				
ensina; em; casa	Artes mecânicas	Ensino	x				
ensina; meninos	Artes mecânicas	Ensino	x				
ensinar; meninos	Artes mecânicas	Ensino	x				
ensinou; gramática	Estudo	Ensino		x			
entalhador	Artes mecânicas	Transformação	x				
ermitão	Igreja	Eclesiástico		x			
ermitoa	Igreja	Eclesiástico		x			
escrevente	Artes mecânicas	Transformação	x				
escritor; genealógico	Artes liberais	Não se sabe		x			
escrivão	Governo da res publica	Administração		x			
escrivão; casa; dízima	Governo da res publica	Administração		x			
escrivão; ad hoc	Igreja	Eclesiástico		x			
escrivão; agravos	Governo da res publica	Administração		x			
escrivão; ajudante	Governo da res publica	Administração		x			
escrivão; alfândega	Governo da res publica	Administração		x			
escrivão; almotaçaria	Governo da res publica	Administração		x			
escrivão; almoxarifado	Governo da res publica	Administração		x			
escrivão; almoxarife	Governo da res publica	Administração		x			
escrivão; apelação; crime	Governo da res publica	Administração		x			
escrivão; aposentadoria	Governo da res publica	Administração		x			
escrivão; apostólico	Igreja	Administração		x			
escrivão; aqueduto	Governo da res publica	Administração	x				
escrivão; arceprestado	Governo da res publica	Administração		x			
escrivão; armas	Governo da res publica	Administração		x			
escrivão; auditoria	Governo da res publica	Administração		x			
escrivão; balança	Governo da res publica	Administração		x			
escrivão; cabido	Governo da res publica	Administração		x			
escrivão; câmara	Governo da res publica	Administração		x			

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe



Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
Escrivão; câmara; arcebispo	Governo da res publica	Administração		x			
escrivão; câmara; eclesiástica	Governo da res publica	Administração		x			
escrivão; câmara; episcopal	Governo da res publica	Administração		x			
escrivão; câmara; mesa	Governo da res publica	Administração		x			
escrivão; casamentos	Governo da res publica	Administração		x			
escrivão; chanceler	Governo da res publica	Administração		x			
escrivão; cível	Governo da res publica	Administração		x			
escrivão; comarca	Governo da res publica	Administração		x			
escrivão; comissário	Igreja	Eclesiástico		x			
escrivão; conservatória	Governo da res publica	Administração		x			
escrivão; contas	Governo da res publica	Administração		x			
escrivão; contos	Governo da res publica	Administração		x			
escrivão; corregedor	Governo da res publica	Administração		x			
escrivão; correição	Governo da res publica	Administração		x			
escrivão; correio; nau	Artes mecânicas	Transformação	x				
escrivão; coutos	Governo da res publica	Administração		x			
escrivão; crime	Governo da res publica	Administração		x			
escrivão; décimas	Governo da res publica	Administração		x			
escrivão; deligências	Governo da res publica	Administração		x			
escrivão; direitos	Governo da res publica	Administração		x			
escrivão; eclesiástico	Governo da res publica	Administração		x			
escrivão; entradas; vinhos	Artes mecânicas	Transformação	x				
escrivão; fazenda	Governo da res publica	Administração		x			
escrivão; feitoria	Governo da res publica	Administração		x			
escrivão; feitos	Governo da res publica	Administração		x			
escrivão; fisco	Governo da res publica	Administração		x			
escrivão; fortificações	Governo da res publica	Administração		x			
escrivão; geral	Governo da res publica	Administração		x			
escrivão; habilitação; oc	Não se sabe	Não se sabe		x			
escrivão; judicial	Artes mecânicas	Administração	x				
escrivão; livros; findos	Governo da res publica	Administração		x			
escrivão; mampostaria	Governo da res publica	Administração		x			
escrivão; matrícula	Governo da res publica	Administração		x			
escrivão; mc	Governo da res publica	Administração		x			
escrivão; meirinho	Governo da res publica	Administração	x				
escrivão; misericórdia	Governo da res publica	Administração		x			
escrivão; mitra	Governo da res publica	Administração		x			
escrivão; mosteiro	Governo da res publica	Administração		x			
escrivão; navio	Governo da res publica	Administração	x				
escrivão; negócios	Artes mecânicas	Transformação	x				
escrivão; obras	Artes mecânicas	Transformação	x				
escrivão; órfãos	Governo da res publica	Administração		x			
escrivão; ouvidoria	Governo da res publica	Administração		x			
escrivão; proprietário	Governo da res publica	Administração		x			
escrivão; provedoria	Governo da res publica	Administração		x			
escrivão; público	Artes mecânicas	Transformação	x				
escrivão; pureza	Governo da res publica	Administração		x			
escrivão; real	Governo da res publica	Administração		x			
escrivão; receita; despesa	Governo da res publica	Administração		x			
escrivão; relação	Governo da res publica	Administração		x			
escrivão; resíduos	Governo da res publica	Administração		x			
escrivão; sé	Igreja	Administração		x			
escrivão; sisas	Governo da res publica	Administração		x			
escrivão; tabaco	Governo da res publica	Administração		x			
escrivão; távola	Governo da res publica	Administração		x			
escrivão; vara	Igreja	Administração		x			
escrivão; vigararia	Igreja	Administração		x			
escrivão; vigário	Igreja	Administração		x			
escrivão; visita; naus	Governo da res publica	Administração		x			

Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
escrivão; visitação	Governo da res publica	Administração		x			
escudeiro	Governo da casa	Doméstico	x				
escultor	Belas artes	Belas artes	x				
esmoler	Governo da casa	Administração		x			
esmoler-mor	Governo da casa	Administração		x			
espadeiro	Artes mecânicas	Transformação	x				
esparteiro	Artes mecânicas	Transformação	x				
espingardas	Artes mecânicas	Transformação	x				
espingardeiro	Artes mecânicas	Transformação	x				
estafeta	Artes mecânicas	Transporte	x				
estalajadeiro	Artes mecânicas	Transformação	x				
estaqueiro-oficial	Artes mecânicas	Transformação	x				
estaqueira	Mercancia	Negócio	x				
estaqueiro	Mercancia	Negócio	x				
esteireiro	Artes mecânicas	Transformação	x				
estribeiro	Governo da casa	Doméstico	x				
estribeiro; menor	Governo da casa	Doméstico	x				
estribeiro; mor	Governo da casa	Doméstico		x			
estudante	Estudo	Ensino		x			
estudante; canonista	Estudo	Ensino		x			
estudante; colegial	Estudo	Ensino		x			
estudante; cursante	Estudo	Ensino		x			
estudante; filósofo	Estudo	Ensino		x			
estudante; matriculado	Estudo	Ensino		x			
estudante; opositor	Estudo	Ensino		x			
estudante; provido	Estudo	Ensino		x			
estudante; teólogo	Estudo	Ensino		x			
estudante; universitário	Estudo	Ensino		x			
estudantes	Estudo	Ensino		x			
estudou	Estudo	Ensino		x			
ex; leitor	Igreja	Eclesiástico		x			
examinador	Igreja	Eclesiástico		x			
examinador; sinodal	Igreja	Eclesiástico		x			
ex-corista	Igreja	Eclesiástico		x			
ex-custódio	Igreja	Eclesiástico		x			
ex-definidor	Igreja	Eclesiástico		x			
executor	Governo da res publica	Administração	x				
executor; mitra	Governo da res publica	Administração	x				
exercício; literário	Artes liberais	Belas artes		x			
ex-leitor	Igreja	Eclesiástico		x			
ex-provincial	Igreja	Eclesiástico		x			
fabrica; embarcações	Artes mecânicas	Transformação	x				
fabricante	Artes mecânicas	Transformação	x				
faz; cabeleiras	Artes mecânicas	Transformação	x				
faz; cestos	Artes mecânicas	Transformação	x				
faz; liteiras	Artes mecânicas	Transformação	x				
faz; viagens	Mercancia	Transporte	x				
fazendeiro	Lavoura	Agricultura	x				
fazer; imagens; santos	Artes mecânicas	Transformação	x				
fazer; panos	Artes mecânicas	Transformação	x				
fazia; louça	Artes mecânicas	Transformação	x				
feitor	Governo da casa	Administração	x				
feitor; aguardentes	Governo da res publica	Administração	x				
feitor; alfândega	Governo da res publica	Administração	x				
feitor; mosteiro	Governo da casa	Administração		x			
feitor; nau	Governo da res publica	Administração	x				
feitor; pescado	Governo da res publica	Administração	x				
feitor; pimenta	Governo da res publica	Administração		x			
feitor-mor	Governo da res publica	Administração		x			

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe

Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
ferrador	Artes mecânicas	Transformação	x				
ferreiro	Artes mecânicas	Transformação	x				
fiandeira	Artes mecânicas	Transformação	x				
fidalgo	Sem ocupação/nobre	Sem ocupação/nobre			x		
fidalgo; escudeiro	Sem ocupação/nobre	Sem ocupação/nobre			x		
filósofo	Não se sabe	Não se sabe		x			
físico; cárceres	Artes liberais	Saúde		x			
físico; so	Artes liberais	Saúde		x			
fiteiro	Artes mecânicas	Transformação	x				
forneira	Artes mecânicas	Transformação	x				
forneiro	Artes mecânicas	Transformação	x				
frade	Igreja	Eclesiástico		x			
frade; menor	Igreja	Eclesiástico		x			
franciscano	Igreja	Eclesiástico		x			
frei	Igreja	Eclesiástico		x			
freira	Igreja	Eclesiástico		x			
freire	Igreja	Eclesiástico		x			
freire; militar	Igreja	Eclesiástico		x			
freire; clérigo	Igreja	Eclesiástico		x			
freire; conventual	Igreja	Eclesiástico		x			
freire; professo	Igreja	Eclesiástico		x			
frequenta; aulas	Estudo	Ensino		x			
fundidor	Artes mecânicas	Transformação	x				
funileiro	Artes mecânicas	Transformação	x				
furriel	Milícia	Defesa		x			
furriel-mor	Milícia	Defesa		x			
genealogista	Artes liberais	Não se sabe		x			
general	Milícia	Defesa		x			
gentil; homem	Governo da casa	Doméstico	x				
geral; da; ordem	Igreja	Eclesiástico		x			
governa; casa	Governo da casa	Administração		x			
governador	Governo da res publica	Administração		x			
governador; bispado	Igreja	Administração		x			
governador; capitão	Governo da res publica	Administração		x			
governador; geral	Governo da res publica	Administração		x			
gravador; diamantes	Artes mecânicas	Transformação	x				
guadamecileiro	Artes mecânicas	Transformação	x				
guarda	Governo da casa	Segurança	x				
guarda; alfândega	Governo da res publica	Segurança	x				
guarda; bens; igreja	Governo da casa	Segurança	x				
guarda; câmara	Governo da res publica	Segurança	x				
guarda; cárceres	Governo da res publica	Segurança	x				
guarda; cartório	Governo da res publica	Segurança	x				
guarda; décimas	Governo da res publica	Segurança	x				
guarda; estudantes	Governo da res publica	Segurança	x				
guarda; estudo	Governo da casa	Segurança	x				
guarda; padres	Governo da casa	Segurança	x				
guarda; pátio	Governo da casa	Segurança	x				
guarda; relação	Governo da res publica	Segurança	x				
guarda; so	Governo da res publica	Segurança	x				
guarda; universidade	Governo da res publica	Segurança	x				
guarda-mor	Governo da res publica	Administração		x			
guarda-mor; saude	Governo da res publica	Saúde		x			
guardar; gado	Lavoura	Pecuária	x				
guardião	Igreja	Eclesiástico		x			
habilitado; desembargo; paço; serviço	Governo da res publica	Administração		x			
hábito; descoberto	Igreja	Eclesiástico		x			
harpista	Belas artes	Belas artes	x				

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe



Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
homem; contrato	Mercancia	Negócio	x				
homem; mar	Navegação	Navegação	x				
homem; marítimo	Navegação	Navegação	x				
homem; meirinho	Governo da res publica	Justiça	x				
homem; negócio	Mercancia	Negócio		x			
homem; saco	Não se sabe	Não se sabe	x				
homem; sem; ofício	Sem ocupação	Sem ocupação	x				
homem; sem; trato; algum	Sem ocupação	Sem ocupação	x				
homem; trabalhador	Lavoura	Agricultura	x				
homem; vara	Governo da res publica	Justiça	x				
homens; saco	Não se sabe	Não se sabe	x				
hortelã	Lavoura	Agricultura	x				
hortelão	Lavoura	Agricultura	x				
hospedeiro; religiosas	Governo da casa	Doméstico	x				
hospitaleiro	Governo da res publica	Administração	x				
imaginário	Artes mecânicas	Transformação	x				
impressor	Artes mecânicas	Transformação	x				
inquiridor	Governo da res publica	Justiça		x			
inquiridor; distribuidor	Governo da res publica	Justiça		x			
inquiridor; proprietário	Governo da res publica	Justiça		x			
inquisidor	Governo da res publica	Administração		x			
intérprete; justiça	Governo da res publica	Justiça	x				
irmão; 3º	Igreja	Eclesiástico		x			
irmão; andador	Igreja	Eclesiástico	x				
irmão; terceiro	Igreja	Eclesiástico		x			
jesuíta	Igreja	Eclesiástico		x			
joalheiro	Artes mecânicas	Transformação	x				
jornal; trabalho; roca	Artes mecânicas	Transformação	x				
jornaleiro	Lavoura	Agricultura	x				
juiz	Governo da res publica	Administração		x			
juiz; adjutor	Igreja	Administração		x			
juiz; alfândega	Governo da res publica	Administração		x			
juiz; aposentadoria	Governo da res publica	Administração		a			
juiz; apostólico	Governo da res publica	Administração		x			
juiz; casa; real	Governo da res publica	Administração		x			
juiz; casamentos	Governo da res publica	Administração		x			
juiz; conservador	Governo da res publica	Administração		x			
juiz; crime	Governo da res publica	Justiça		x			
juiz; degredados	Governo da res publica	Justiça		x			
juiz; direitos; reais	Governo da res publica	Administração		x			
juiz; executor	Governo da res publica	Administração		x			
juiz; feitos; coroa	Governo da res publica	Administração		x			
juiz; fisco	Governo da res publica	Administração		x			
juiz; fora	Governo da res publica	Justiça		x			
juiz; freires	Governo da res publica	Administração		x			
juiz; geral; ordens	Governo da res publica	Administração		x			
juiz; habilitações	Governo da res publica	Administração		x			
juiz; honorário	Governo da res publica	Administração		x			
juiz; justificações	Governo da res publica	Administração		x			
juiz; mercador	Governo da res publica	Justiça		x			
juiz; ordem	Igreja	Administração		x			
juiz; ordenança	Governo da res publica	Administração		x			
juiz; ordinário	Governo da res publica	Administração		x			
juiz; órfãos	Governo da res publica	Administração		x			
juiz; povo	Artes mecânicas	Transformação	x				
juiz; prata	Governo da res publica	Administração		x			
juiz; presídio	Governo da res publica	Justiça		x			
juiz; rendas	Governo da res publica	Administração		x			
juiz; resíduos	Governo da res publica	Administração		x			

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe



Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
juiz; sete; varas	Governo da res publica	Administração		x			
juiz; sinodal	Governo da res publica	Administração		x			
juiz; tombo	Governo da res publica	Administração		x			
Juiz; vintena	Governo da res publica	Administração	x				
juiz-mor	Governo da res publica	Administração		x			
lacerantes; pedra	Artes mecânicas	Transformação	x				
ladrilhador	Artes mecânicas	Transformação	x				
lanterneiro	Artes mecânicas	Transformação	x				
lapidário	Artes mecânicas	Transformação	x				
laranjeiro	Lavoura	Agricultura	x				
latoeiro	Artes mecânicas	Transformação	x				
lavadeira	Artes mecânicas	Transformação	x				
lavradeira	Lavoura	Agricultura	x				
lavrador	Lavoura	Agricultura	x				x
lavrador; abastado	Lavoura	Agricultura		x			
lavrador; honrado	Lavoura	Agricultura		x			
lavradora	Lavoura	Agricultura	x				x
lavrante; pedra	Artes mecânicas	Transformação	x				
leitor	Igreja	Ensino		x			
leitor; jubilado	Igreja	Ensino		x			
lente	Estudo	Ensino		x			
levar; cartas; correio	Artes mecânicas	Transporte	x				
licenciado	Artes liberais	Não se sabe		x			
livreiro	Artes mecânicas	Transformação	x				
loja; marceria	Mercancia	Negócio	x				
lojeira	Mercancia	Negócio	x				
maioral; carretas	Artes mecânicas	Transporte	x				
maioral; carretas	Artes mecânicas	Transporte	x				
mamosteiro	Governo da res publica	Administração		x			
mamosteiro; menor	Governo da res publica	Administração		x			
mamosteiro-mor	Governo da res publica	Administração		x			
marceiro	Mercancia	Negócio	x				
marceneiro	Artes mecânicas	Transformação	x				
marchante	Mercancia	Negócio	x				
mareante	Navegação	Navegação	x				
marinheiro	Navegação	Navegação	x				
médico	Não se sabe	Saúde				x	x
médico; aprovado; universidade	Artes liberais	Saúde		x			
médico; assistente	Artes liberais	Saúde		x			x
médico; formado	Artes liberais	Saúde		x			
médico; graduado	Artes liberais	Saúde		x			
médico; partidário; universidade	Artes liberais	Saúde		x			
médico; partidista; universidade	Artes liberais	Saúde		x			
médico; rainha	Artes liberais	Saúde		x			
médico; real; câmara	Artes liberais	Saúde		x			
médico; rei	Artes liberais	Saúde		x			
medideira	Artes mecânicas	Transformação	x				
medidor	Artes mecânicas	Transformação	x				
meio; cônego	Igreja	Eclesiástico		x			
meio; prebendado	Igreja	Eclesiástico		x			
meio-cônego	Igreja	Eclesiástico		x			
meirinho	Governo da res publica	Justiça	x				
membro; casa; suplicação	Governo da res publica	Administração		x			
membro; cg	Governo da res publica	Administração		x			
membro; conselho; geral	Governo da res publica	Administração		x			
membro; conselho; portugal; madrid	Governo da res publica	Administração		x			

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe

Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
membro; conselho; ultramarino	Governo da res publica	Administração		x			
membro; junta; estados	Governo da res publica	Administração		x			x
mendigo	Sem ocupação	Sem ocupação	x				
meneava	Mercancia	Negócio	x				
mercador	Mercancia	Negócio	x				
mercador; capela	Mercancia	Negócio	x				
merceeira	Não se sabe	Não se sabe	x				x
meretriz	Mercancia	Negócio	x				
mesário	Não se sabe	Não se sabe		x			
mesteiral	Artes mecânicas	Transformação	x				
mester	Artes mecânicas	Transformação	x				
mestre; caixeiro	Mercancia	Negócio	x				
mestre; campo	Milícia	Defesa		x			
mestre; capela	Igreja	Eclesiástico		x			
mestre; capela; música	Igreja	Eclesiástico		x			
mestre; capitão; nau	Navegação	Navegação		x			
mestre; capitão; navio	Navegação	Navegação		x			
mestre; caravela	Artes mecânicas	Transporte	x				
mestre; carreira	Artes mecânicas	Transporte	x				
mestre; cerimónias	Igreja	Eclesiástico		x			
mestre; cerimónias; sé	Igreja	Eclesiástico		x			
mestre; cerimónias; universidade	Governo da res publica	Administração		x			
mestre; colégio	Artes liberais	Ensino		x			
mestre; embarcação	Artes mecânicas	Transporte	x				
mestre; escola	Igreja	Eclesiástico		x			
mestre; escola; sé	Igreja	Eclesiástico		x			
mestre; estanco	Mercancia	Negócio	x				
mestre; estudantes	Artes liberais	Ensino		x			x
mestre; galeão	Artes mecânicas	Transporte	x				
mestre; gramática	Artes liberais	Ensino		x			x
mestre; meninos	Artes mecânicas	Ensino	x				
mestre; menores	Artes mecânicas	Ensino	x				
mestre; na; escola	Artes liberais	Ensino		x			
mestre; nau	Artes mecânicas	Transporte	x				
mestre; navio	Artes mecânicas	Transporte	x				
mestre; noviços	Igreja	Eclesiástico		x			
mestre; obras	Artes mecânicas	Transformação	x				
mestre; ordem	Igreja	Eclesiástico		x			
mestre; piloto	Navegação	Navegação	x				
mestre; postas; rei	Artes mecânicas	Transporte	x				
mestre; reparações	Igreja	Eclesiástico		x			
mestre; senhorio; embarcação	Navegação	Navegação		x			
mestre; senhorio; nau	Navegação	Navegação		x			
mestre; senhorio; navio	Navegação	Navegação					
mestre; serviços; convento; serra	Igreja	Eclesiástico		x			
mestre; sua; província	Igreja	Eclesiástico		x			
mestre; sua; tenda	Artes mecânicas	Transformação	x				
mestre; tendas	Artes mecânicas	Transformação	x				
mestre; teologia	Igreja	Eclesiástico		x			
mestre; terceiros	Igreja	Eclesiástico		x			
mestre-escola	Igreja	Eclesiástico		x			
militante; conquistas; orientais	Milícia	Defesa	x				
ministro; letras	Governo da res publica	Administração		x			
ministro; relação	Governo da res publica	Justiça		x			

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe

Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
ministro; sacristia	Igreja	Eclesiástico		x			
ministro; so	Governo da res publica	Administração		x			
missionário	Igreja	Eclesiástico		x			
moça; convento	Governo da casa	Doméstico	x				
moça; coro	Igreja	Eclesiástico		x			
moço; almocreve	Governo da casa	Doméstico	x				
moço; casa; arreios	Governo da casa	Doméstico	x				
moço; estribeira	Governo da casa	Doméstico	x				
moço; guarda-roupa	Governo da casa	Doméstico	x				
moço; saco	Não se sabe	Não se sabe	x				x
moço; servir	Governo da casa	Doméstico	x				
moço; soldada	Governo da casa	Doméstico	x				
moleiro	Artes mecânicas	Transformação	x				
monge	Igreja	Eclesiástico		x			
monteiro-mor	Governo da casa	Administração		x			
mordomo	Igreja	Eclesiástico		x			
músico	Belas artes	Belas artes				x	x
não; tem; ocupação	Sem ocupação	Sem ocupação				x	x
não; tem; ofício	Sem ocupação	Sem ocupação				x	x
navegador	Navegação	Navegação	x				
negociante	Mercancia	Negócio		x			
notário	Governo da res publica	Administração	x				
notário; apostólico	Governo da res publica	Administração		x			
notário; cg	Governo da res publica	Administração		x			
notário; conselho; geral	Governo da res publica	Administração		x			
notário; extranumerário	Governo da res publica	Administração	x				
notário; ic	Governo da res publica	Administração		x			
notário; ie	Governo da res publica	Administração		x			
notário; il	Governo da res publica	Administração		x			
notário; público	Governo da res publica	Administração	x				
notário; secreto	Governo da res publica	Administração		x			
notário; so	Governo da res publica	Administração		x			
notário; supra; numerário	Governo da res publica	Administração	x				
notário; supranumerário	Governo da res publica	Administração	x				
notário; tribunal	Governo da res publica	Administração		x			
notário; visitação	Governo da res publica	Administração		x			
noviço	Igreja	Eclesiástico		x			
número; serviço; casa	Governo da casa	Doméstico	x				
odreiro	Artes mecânicas	Transformação	x				
oficial; alveitar	Artes mecânicas	Transformação	x				
oficial; auditório; eclesiástico	Governo da res publica	Administração	x				
oficial; conselho; fazenda	Governo da res publica	Administração	x				
oficial; contínuo	Governo da res publica	Administração	x				
oficial; fisco	Governo da res publica	Administração	x				
oficial; juízo eclesiástico	Governo da res publica	Administração	x				
oficial; justiça	Governo da res publica	Justiça		x			x
oficial; maior	Governo da res publica	Administração		x			
oficial; pagador	Governo da res publica	Administração	x				
oficial; secretaria; conselho; ultramarino	Governo da res publica	Administração		x			
oficial; secretaria; estado	Governo da res publica	Administração		x			
oficial; vedoria; exército; elvas	Governo da res publica	Administração		x			
oficial-maior	Governo da res publica	Administração		x			
ofício; judicial	Governo da res publica	Justiça				x	
oleiro	Artes mecânicas	Transformação	x				
opositor	Estudo	Ensino		x			
opositor; cadeira	Estudo	Ensino		x			
opositor; cânones	Estudo	Ensino		x			

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe

Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
opositor; condutário	Estudo	Ensino		x			
opositor; lugares; letras	Governo da res publica	Administração		x			
opositor; universidade	Estudo	Ensino		x			
ordenado; ordens; sacras	Igreja	Eclesiástico		x			
ordinário; capela; casa; real	Igreja	Eclesiástico			x		
organista	Belas artes	Belas artes				x	x
ourives	Artes mecânicas	Transformação	x				
ouvidor	Governo da res publica	Administração		x			
padeira	Artes mecânicas	Transformação	x				
padeiro	Artes mecânicas	Transformação	x				
padre	Igreja	Eclesiástico		x			
pagador; geral	Governo da res publica	Administração		x			
pajem	Governo da casa	Doméstico	x				
palmeiro	Artes mecânicas	Transformação	x				
palmeira	Artes mecânicas	Transformação	x				
palmeiro	Artes mecânicas	Transformação	x				
pantufeiro	Artes mecânicas	Transformação	x				
pároco	Igreja	Eclesiástico		x			
parteira	Artes mecânicas	Transformação	x				
partidista	Artes liberais	Saúde		x			
partidor	Artes mecânicas	Administração	x				
pasteleiro	Artes mecânicas	Transformação	x				
pastor	Lavoura	Pecuária	x				
patrão-mor	Artes mecânicas	Administração	x				
pedreiro	Artes mecânicas	Transformação	x				
peneireiro	Artes mecânicas	Transformação	x				
penitenciário; sé	Igreja	Eclesiástico		x			
pensionário	Não se sabe	Não se sabe		x			
penteiro	Artes mecânicas	Transformação	x				
perdentário; apostólico	Igreja	Administração		x			
pescadeira	Artes mecânicas	Pesca	x				
pescador	Artes mecânicas	Pesca	x				
picadeiro	Artes mecânicas	Transformação	x				
picheiro	Artes mecânicas	Transformação	x				
piloto	Navegação	Navegação	x				
pintor	Artes mecânicas	Transformação	x				x
pintor; louça	Artes mecânicas	Transformação	x				x
pobre	Sem ocupação	Sem ocupação	x				
pojeiro	Artes mecânicas	Transformação	x				
polieiro; prata	Artes mecânicas	Transformação	x				
pomareiro	Lavoura	Agricultura	x				
porcionista	Estudo	Ensino		x			
porteiro	Governo da res publica	Administração	x				
prateiras	Artes mecânicas	Transformação	x				
praticante	Governo da res publica	Ensino	x				
prebendeiro	Mercancia	Negócio	x				
pregador	Igreja	Eclesiástico		x			
prelado	Igreja	Eclesiástico			x		
prepósito	Igreja	Eclesiástico		x			
presbítero	Igreja	Eclesiástico		x			
presidente	Governo da res publica	Administração			x		
presidente; hospício	Governo da res publica	Administração		x			
presilhão	Artes mecânicas	Transformação	x				
primeira; cadeira; il	Governo da res publica	Administração		x			
primeiro; boticário	Artes mecânicas	Saúde	x				
primeiro; substituto; universidade	Estudo	Ensino		x			
primeiro; vigário	Igreja	Eclesiástico		x			
prior	Igreja	Eclesiástico		x			

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe



Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
prioresa	Igreja	Eclesiástico		x			
prior-mor	Igreja	Eclesiástico		x			
prioste	Mercancia	Negócio	x				x
procurador	Governo da casa	Administração					x
procurador; bispo	Igreja	Administração		x			
procurador; câmara	Governo da res publica	Administração		x			
procurador; capelas; coroa	Governo da res publica	Administração		x			
procurador; cárceres	Governo da res publica	Administração		x			
procurador; causas	Governo da casa	Administração	x				
procurador; cidade	Governo da res publica	Administração		x			
procurador; comarca	Governo da res publica	Administração		x			
procurador; concelho	Governo da res publica	Administração					x
procurador; cortes	Governo da res publica	Administração		x			
procurador; fazenda; real	Governo da res publica	Administração		x			
procurador; fiscal	Governo da res publica	Administração		x			
procurador; fisco	Governo da res publica	Administração		x			
procurador; freiras	Governo da casa	Administração	x				
procurador; geral	Governo da res publica	Administração		x			
procurador; ic	Governo da res publica	Administração		x			
procurador; ie	Governo da res publica	Administração		x			
procurador; il	Governo da res publica	Administração		x			
procurador; inquisições	Governo da res publica	Administração		x			
procurador; letrado	Governo da res publica	Administração		x			
procurador; misericórdia	Governo da casa	Administração	x				
procurador; mitra	Governo da casa	Administração	x				
procurador; número	Governo da res publica	Administração		x			
procurador; presos	Governo da res publica	Administração		x			
procurador; religiosas	Governo da casa	Administração	x				
procurador; senado	Governo da res publica	Administração		x			
procurador; so	Governo da res publica	Administração		x			
procurador-geral	Governo da res publica	Administração		x			
professa; ordem	Igreja	Eclesiástico		x			
professo; ordem	Igreja	Eclesiástico		x			
professor	Igreja	Eclesiástico		x			
professou	Estudo	Ensino		x			
promotor	Governo da res publica	Administração		x			
promotor; fiscal	Governo da res publica	Administração		x			
promotor; so	Governo da res publica	Administração		x			
propriedade; ofício; despenseiro	Governo da res publica	Administração		x			
propriedade; ofício; meirinho	Governo da res publica	Administração		x			
proprietária; ofício; despenseiro	Governo da res publica	Administração		x			
proprietária; ofício; solicitador	Governo da res publica	Administração		x			
proprietário; casa; pasto; bilhar	Mercancia	Negócio	x				
proprietário; engenhos; açúcar	Lavoura	Transformação		x			
proprietário; ofício; advogado; auditórios	Governo da res publica	Administração		x			
proprietário; ofício; aferidor	Governo da res publica	Administração		x			
proprietário; ofício; apostólico	Governo da res publica	Administração		x			
proprietário; ofício; contador; auditórios	Governo da res publica	Administração		x			
proprietário; ofício; correio; postas	Governo da res publica	Transporte		x			
proprietário; ofício; corretor; fazendas	Mercancia	Negócio		x			
proprietário; ofício; despenseiro	Governo da res publica	Administração		x			

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe

Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
proprietário; ofício; distribuidor; auditórios	Governo da res publica	Administração		x			
proprietário; ofício; escrivão; correição	Governo da res publica	Administração		x			
proprietário; ofício; escrivão; fisco	Governo da res publica	Administração		x			
proprietário; ofício; escrivão; judicial	Governo da res publica	Administração		x			
proprietário; ofício; escrivão; provedoria	Governo da res publica	Administração		x			
proprietário; ofício; inquiridor; auditórios	Governo da res publica	Administração		x			
proprietário; ofício; juiz; alfândega	Governo da res publica	Administração		x			
proprietário; ofício; juiz; órfãos	Governo da res publica	Administração		x			
proprietário; ofício; tabelião; geral	Governo da res publica	Administração		x			
proprietário; ofício; tabelião; judicial	Governo da res publica	Administração		x			
proprietário; ofício; tabelião; notas	Governo da res publica	Administração		x			
proprietário; ofícios; contador	Governo da res publica	Administração		x			
proprietário; ofícios; corretor; fazendas	Governo da res publica	Administração		x			
proprietário; ofícios; escrivão; órfãos	Governo da res publica	Administração		x			
proprietário; ofícios; tabelião; notas	Governo da res publica	Administração		x			
proprietário; vara; ouvidoria; Universidade	Governo da res publica	Administração		x			
proto; notário	Governo da res publica	Administração		x			
proto-notário	Governo da res publica	Administração		x			
protonotário; apostólico	Governo da res publica	Administração		x			
provedor	Governo da res publica	Administração		x			
provedor; misericórdia	Governo da casa	Administração		x			
provedor; obras	Governo da res publica	Administração		x			
provedor; togado	Governo da res publica	Administração		x			
provedor-mor	Governo da res publica	Saúde		x			
provido; igreja	Igreja	Eclesiástico		x			
provincial	Igreja	Eclesiástico		x			
provisor	Igreja	Eclesiástico		x			
público; notário	Governo da res publica	Administração	x				
público; tabelião	Governo da res publica	Administração	x				
pupilo	Estudo	Ensino		x			
qualificador	Igreja	Eclesiástico		x			
qualificador; revedor	Igreja	Eclesiástico		x			
qualificador; so	Igreja	Eclesiástico		x			
quaternário	Igreja	Eclesiástico		x			
quintador	Governo da res publica	Administração	x				
quintaneiro	Lavoura	Agricultura	x				
ratificador; so	Governo da res publica	Administração		x			
ratificante	Governo da res publica	Administração		x			
recebe; ordens; missa	Igreja	Eclesiástico		x			
recebedor	Governo da res publica	Administração	x				
recebedor-geral	Governo da res publica	Administração	x				
recolhida	Igreja	Ensino		x			
recondução; serventia	Governo da res publica	Administração	x				x
recoveiro	Artes mecânicas	Transporte	x				
recoveiro; so	Artes mecânicas	Transporte	x				

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe

Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
reformador; congregação	Igreja	Eclesiástico		x			
reformador; ordens	Igreja	Eclesiástico		x			
reformador; universidade	Governo da res publica	Ensino		x			
regatão	Mercancia	Negócio	x				
regateira	Mercancia	Negócio	x				
regateiro	Mercancia	Negócio	x				
regedor	Não se sabe	Não se sabe				x	
regedor; justiça	Governo da res publica	Justiça		x			
regente; estudo	Estudo	Ensino		x			
reitor	Igreja	Eclesiástico		x			
reitor; colado	Igreja	Eclesiástico		x			
reitor; colégio	Igreja	Eclesiástico		x			
reitor; pensionário	Igreja	Eclesiástico		x			
reitor; perpétuo	Igreja	Eclesiástico		x			
reitor; reservatário	Igreja	Eclesiástico		x			
reitor; universidade	Governo da res publica	Administração		x			
religiosa	Igreja	Eclesiástico		x			
religiosas	Igreja	Eclesiástico		x			
religioso	Igreja	Eclesiástico		x			
religioso; leigo	Igreja	Eclesiástico		x			
relojoeiro	Artes mecânicas	Transformação	x				
rendeiro	Lavoura	Agricultura	x				
reposteiro	Governo da casa	Doméstico	x				
representou; cg	Governo da res publica	Administração		x			
requerente	Governo da casa	Administração	x				
requerente; causas	Governo da casa	Administração	x				
requerente; misericórdia	Governo da casa	Administração	x				
reservatário	Igreja	Eclesiástico		x			
responsável; colégio; órfãos	Governo da res publica	Administração		x			
responsável; papéis; secretaria; estado	Governo da res publica	Administração		x			
retroseiro	Artes mecânicas	Transformação	x				
revedor; so	Igreja	Eclesiástico		x			
roupavilheiro	Mercancia	Negócio	x				
saboeira	Artes mecânicas	Transformação	x				
saboeiro	Artes mecânicas	Transformação	x				
sacerdote	Igreja	Eclesiástico		x			
sacristão	Igreja	Administração	x				
sangrador	Artes mecânicas	Saúde	x				
sangrador; so	Artes mecânicas	Saúde	x				
sapateiro	Artes mecânicas	Transformação	x				
sargento	Milícia	Defesa		x			
sargento-mor	Milícia	Defesa		x			
seareiro	Lavoura	Agricultura	x				
secretário	Governo da casa	Administração	x				
secretário; câmara	Governo da res publica	Administração					x
secretário; câmara; eclesiástica	Governo da res publica	Administração		x			
secretário; bispo	Governo da casa	Administração		x			
secretário; casa; despacho	Governo da res publica	Administração		x			
secretário; causas; so; corte	Governo da res publica	Administração		x			
secretário; causas; so; portugal	Governo da res publica	Administração		x			
secretário; cg	Governo da res publica	Administração		x			
secretário; chancelaria	Governo da res publica	Administração		x			
secretário; conselho; geral	Governo da res publica	Administração		x			
secretário; custódia	Igreja	Administração		x			
secretário; estado	Governo da res publica	Administração		x			
secretário; exército	Governo da res publica	Administração		x			
secretário; geral; so	Governo da res publica	Administração		x			

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe



Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
secretário; igreja	Igreja	Eclesiástico		x			
secretário; inquisição	Governo da res publica	Administração		x			
secretário; ligado; so	Governo da res publica	Administração		x			
secretário; mercês	Governo da res publica	Administração		x			
secretário; mesa; consciência	Governo da res publica	Administração		x			
secretário; rei	Governo da res publica	Administração		x			
secretário; so	Governo da res publica	Administração		x			
secretário; universidade	Governo da res publica	Administração		x			
selador; alfândega	Artes mecânicas	Administração	x				
seleiro	Artes mecânicas	Transformação	x				
sem; ocupação	Sem ocupação	Sem ocupação					x
sem; ofício	Não se sabe	Não se sabe					x
semeia; ferragial	Lavoura	Agricultura	x				
senador; cidade	Governo da res publica	Administração		x			
senhorio; nau	Navegação	Navegação		x			
sergueiro	Artes mecânicas	Transformação	x				
serrador	Artes mecânicas	Transformação	x				
serralheiro	Artes mecânicas	Transformação	x				
serve; majestade	Governo da res publica	Não se sabe		x			x
serve; cárceres; inquisição	Governo da res publica	Administração	x				
serve; cargos; misericórdia	Governo da res publica	Administração		x			
serve; coro	Igreja	Eclesiástico		x			
serve; depositário	Governo da res publica	Administração		x			
serve; dispenseiro	Governo da res publica	Administração	x				
serve; escrevente	Governo da res publica	Administração	x				
serve; escrivão	Governo da res publica	Administração					x
serve; escrivão; diligências	Governo da res publica	Administração		x			
serve; escrivão; visita	Governo da res publica	Administração		x			
serve; igreja	Não se sabe	Não se sabe				x	
serve; inquisidor	Governo da res publica	Administração		x			
serve; juiz; vila	Governo da res publica	Administração		x			
serve; juízo; geral	Governo da res publica	Administração		x			
serve; madrid; ordem; inquisidor	Governo da res publica	Administração		x			
serve; majestade; lugares; letras	Governo da res publica	Administração		x			
serve; marinha	Navegação	Defesa		x			x
serve; milícia	Milícia	Defesa		x			
serve; moças	Governo da casa	Doméstico	x				
serve; padres	Governo da casa	Doméstico	x				
serve; partidor; órfãos	Governo da res publica	Administração	x				
serve; prisões; diligências; inquisição	Governo da res publica	Administração		x			
serve; secretário; ao; provisor	Governo da res publica	Administração		x			
serve; so; transporte	Governo da res publica	Transporte		x			
serve; tesoureiro	Governo da res publica	Administração					x
servente	Artes mecânicas	Transformação	x				
serventia; cargo; juiz; órfãos	Governo da res publica	Administração		x			
serventia; escrivão	Governo da res publica	Administração	x				
serventia; guarda; cárceres; so	Governo da res publica	Administração	x				
serventia; meirinho	Governo da res publica	Justiça	x				
serventia; ofício; dispenseiro; so	Governo da res publica	Administração	x				
serventia; ofício; juiz; Coutada	Governo da res publica	Administração		x			
serventia; solicitador; cg	Governo da res publica	Administração	x				
serventuário; alcaide; cárcere	Governo da res publica	Administração	x				

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe



Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
serventuário; escrivão dos órfãos	Governo da res publica	Administração		x			
serventuário; escrivão; almoxarifado	Governo da res publica	Administração	x				
serventuário; meirinho	Governo da res publica	Administração	x				
serventuário; porteiro	Governo da res publica	Administração	x				
serventuário; solicitador	Governo da res publica	Administração	x				
serventuário; tabelião; judicial	Governo da res publica	Administração	x				
servia; cirurgião; cárceres	Artes mecânicas	Saúde	x				
servia; escrivão; so	Governo da res publica	Administração		x			
servia; santo; ofício	Governo da res publica	Administração		x			
servia; so	Governo da res publica	Administração		x			
serviços; guerra; mazagão	Milícia	Defesa		x			
Serviu; cárceres	Governo da res publica	Administração	x				
Serviu; cardeal; dom; henrique	Governo da casa	Doméstico		x			
Serviu; casa; beneficiado	Governo da res publica	Doméstico					x
serviu; conde	Governo da casa	Doméstico					x
Serviu; il	Governo da res publica	Administração		x			
serviu; misericórdia	Governo da res publica	Administração		x			
serviu; ofícios; república	Governo da res publica	Administração		x			
serviu; rei	Milícia	Defesa		x			x
serviu; vara; meirinho	Governo da res publica	Justiça	x				
servo; misericórdia	Governo da casa	Doméstico	x				
sindicante	Governo da casa	Administração		x			
síndico	Governo da casa	Administração		x			
sineiro	Artes mecânicas	Transformação	x				
singaleiro	Artes mecânicas	Transformação	x				
singeleiro	Lavoura	Agricultura	x				
sirgueira	Artes mecânicas	Transformação	x				
sirgueiro	Artes mecânicas	Transformação	x				
siseiro	Governo da res publica	Administração	x				
sócio; loja	Mercancia	Negócio	x				
soldado	Milícia	Defesa	x				
solicitador	Governo da res publica	Administração	x				
solicitador; conselho; geral	Governo da res publica	Administração	x				
solicitador; contínuo	Governo da res publica	Administração	x				
solicitador; fisco	Governo da res publica	Administração	x				
solicitador; so	Governo da res publica	Administração	x				
sombreireiro	Artes mecânicas	Transformação	x				
subchante	Igreja	Eclesiástico		x			
sub-chante	Igreja	Eclesiástico		x			
subdiácono	Igreja	Eclesiástico		x			
sub-diácono	Igreja	Eclesiástico		x			
subscrite; conservador	Não se sabe	Não se sabe				x	
sub-tesoureiro	Governo da res publica	Administração	x				x
superintendente	Governo da res publica	Administração		x			
superintendente; descaminhos	Governo da res publica	Administração		x			
superintendente-geral; alfândega	Governo da res publica	Administração		x			
surrador	Artes mecânicas	Transformação	x				
tabelião	Governo da res publica	Administração	x				
tabelião; apostólico	Governo da res publica	Administração	x				
tabelião; judicial	Governo da res publica	Administração	x				
tabelião; notas	Governo da res publica	Administração	x				
tabelião; público	Governo da res publica	Administração	x				
talheiro	Artes mecânicas	Transformação	x				
tamanqueiro	Artes mecânicas	Transformação	x				
tangedor; charamela	Belas artes	Belas artes	x				
tangedor; sacabuxa	Belas artes	Belas artes	x				

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe

Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
tanoeiro	Artes mecânicas	Transformação	x				
taverneiro	Mercancia	Negócio	x				
tecedeira	Artes mecânicas	Transformação	x				
tecedor; volantes	Artes mecânicas	Transformação	x				
tecelão	Artes mecânicas	Transformação	x				
telheiro	Artes mecânicas	Transformação	x				
tem; loja	Mercancia	Negócio	x				
tem; taberna	Mercancia	Negócio	x				
tem; tenda; linhas	Mercancia	Negócio	x				
tem; tenda; sedas	Mercancia	Negócio	x				
tendeira	Mercancia	Negócio	x				
tendeiro	Mercancia	Negócio	x				
tenente	Milícia	Defesa		x			
tenente; general	Milícia	Defesa		x			
tenente-ajudante	Milícia	Defesa		x			
tenente-coronel	Milícia	Defesa		x			
teólogo	Igreja	Eclesiástico		x			
terceira; carmo	Igreja	Eclesiástico		x			
terceira; francisco	Igreja	Eclesiástico		x			
terceira; hábito	Igreja	Eclesiástico		x			
tercenário	Igreja	Eclesiástico		x			
tesoureiro	Governo da res publica	Administração					x
tesoureiro; executor	Governo da res publica	Administração		x			
tesoureiro; igreja	Igreja	Administração	x				
tesoureiro; matriz	Igreja	Administração	x				
tesoureiro; mc	Governo da res publica	Administração		x			
tesoureiro; so	Governo da res publica	Administração		x			
tesoureiro-geral	Governo da res publica	Administração		x			
tesoureiro-mor	Igreja	Eclesiástico		x			
tesoureiro-real	Governo da res publica	Administração		x			
tintureiro	Artes mecânicas	Transformação	x				
toma; ordens; sacras	Igreja	Eclesiástico		x			
torneiro	Artes mecânicas	Transformação	x				
tosador	Artes mecânicas	Transformação	x				
trabalha; atafona	Artes mecânicas	Transformação	x				
trabalha; tenda	Artes mecânicas	Transformação	x				
trabalhador	Lavoura	Agricultura	x				
trabalhador; corda	Artes mecânicas	Transformação	x				
trabalhador; enxada	Lavoura	Agricultura	x				
trabalhadora	Lavoura	Agricultura	x				
trabalhadores	Lavoura	Agricultura	x				
trabalhar; ceifa	Lavoura	Agricultura	x				
trapeiro	Mercancia	Negócio	x				
trata; cavalgadas	Lavoura	Pecuária	x				
trata; couros	Artes mecânicas	Transformação	x				
trata; merdaes;	Lavoura	Agricultura		x			
tratante	Mercancia	Negócio	x				
trato dos fornos	Artes mecânicas	Transformação	x				
tutor	Governo da casa	Administração		x			x
vagabundo	Sem ocupação	Sem ocupação	x				
vedor; águas	Artes mecânicas	Administração	x				
vedor; câmara	Governo da res publica	Administração	x				
vedor; fazenda	Governo da res publica	Administração		x			
vedor; geral	Governo da res publica	Administração		x			
vedor; justiça	Governo da res publica	Justiça		x			
veleira	Artes mecânicas	Transformação	x				
vende; aguardente	Mercancia	Negócio	x				
vende; chapéus	Mercancia	Negócio	x				
vende; hortaliça	Mercancia	Negócio	x				

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe

Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
vende; tabaco	Mercancia	Negócio	x				
vende; vinhos	Mercancia	Negócio	x				
vendedeira	Mercancia	Negócio	x				
vendeira	Mercancia	Negócio	x				
vendeiro	Mercancia	Negócio	x				
vendeiro; vinhos	Mercancia	Negócio	x				
vendia; pão	Mercancia	Negócio	x				
vereador	Governo da res publica	Administração		x			
vestiário	Artes mecânicas	Administração	x				
vestimenteiro	Artes mecânicas	Transformação	x				
vice-conservador	Governo da res publica	Administração		x			
vice-reitor	Governo da res publica	Administração		x			
vice-vigário	Igreja	Eclesiástico		x			
vidraceiro	Artes mecânicas	Transformação	x				
vidreiro	Artes mecânicas	Transformação	x				
vigário	Igreja	Eclesiástico		x			
vigário; vara	Igreja	Administração		x			
vinhateiro	Lavoura	Agricultura	x				
violeiro	Artes mecânicas	Transformação	x				
visitador	Governo da res publica	Administração		x			
visitador; bispado	Governo da res publica	Administração		x			
visitador; geral	Governo da res publica	Administração		x			
visitador; naus	Governo da res publica	Administração		x			
visitador; ordinário	Governo da res publica	Administração		x			
visitador; so	Governo da res publica	Administração		x			
vive; agência	Mercancia	Negócio		x			
vive; agência; comércio	Mercancia	Negócio		x			
vive; agência; contrato	Mercancia	Negócio		x			
vive; agência; negócio	Mercancia	Negócio		x			
vive; agência; trabalho	Lavoura	Agricultura	x				
vive; agências	Mercancia	Negócio		x			
vive; almofada	Artes mecânicas	Transformação	x				
vive; bens	Viver da sua fazenda	Viver da sua fazenda		x			
vive; bens; fazendas	Viver da sua fazenda	Viver da sua fazenda		x			
vive; capela	Viver da sua fazenda	Viver da sua fazenda		x			
vive; contrato	Mercancia	Negócio		x			
vive; curiosidades	Não se sabe	Não se sabe				x	
vive; fazenda	Viver da sua fazenda	Viver da sua fazenda		x			
vive; fazenda; agência	Viver da sua fazenda	Viver da sua fazenda		x			
vive; fazenda; botica	Artes mecânicas	Transformação	x				
vive; fazenda; contratos	Mercancia	Negócio		x			
vive; fazenda; esteireiro	Artes mecânicas	Transformação	x				
vive; fazenda; fidalgamente	Viver da sua fazenda	Viver da sua fazenda		x			
vive; fazenda; indústria	Mercancia	Negócio		x			
vive; fazenda; morgado	Viver da sua fazenda	Viver da sua fazenda		x			
vive; fazenda; negócio; agência	Mercancia	Negócio		x			
vive; fazenda; negócio; mercância	Mercancia	Negócio		x			
vive; fazenda; pedreiro	Artes mecânicas	Transformação		x			
vive; fazenda; rendas	Viver da sua fazenda	Viver da sua fazenda		x			
vive; fazenda; sapateiro	Artes mecânicas	Transformação	x				
vive; fazenda; sem; ofício	Viver da sua fazenda	Viver da sua fazenda		x			
vive; fazenda; trabalho	Lavoura	Agricultura	x				
vive; fazenda; trata;	Viver da sua fazenda	Viver da sua fazenda		x			
vive; fazenda; trato	Mercancia	Negócio		x			
vive; fazenda; usa; ofício	Viver da sua fazenda	Viver da sua fazenda		x			
vive; fazendas; gados; lavouras	Lavoura	Agricultura		x			
vive; fazendas; governança	Viver da sua fazenda	Viver da sua fazenda		x			

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe

Palavras-chave	Setor de ocupação	Setor de atividade	ME	NM	NO	NS	Depende do contexto
vive; fazendas; granjarias; lavouras	Lavoura	Agricultura		x			
vive; fazendas; lavoura	Lavoura	Agricultura	x				
vive; fazendas; lavrador	Lavoura	Agricultura		x			
vive; fazendas; limpa; abastadamente	Viver da sua fazenda	Viver da sua fazenda		x			
vive; fazendas; morgadio	Viver da sua fazenda	Viver da sua fazenda		x			
vive; fazendas; morgado	Viver da sua fazenda	Viver da sua fazenda		x			
vive; fazendas; negócios	Mercancia	Negócio		x			
vive; fazendas; rendas	Viver da sua fazenda	Viver da sua fazenda		x			
vive; fazendas; rendimentos	Viver da sua fazenda	Viver da sua fazenda		x			
vive; fazendas; vinhas	Lavoura	Agricultura	x				
vive; fiéis	Igreja	Eclesiástico		x			
vive; granjeamento	Lavoura	Agricultura	x				
vive; granjeio	Lavoura	Agricultura	x				
vive; indústria	Lavoura	Agricultura	x				
vive; indústria; trabalho	Lavoura	Agricultura	x				
vive; jornais	Lavoura	Agricultura	x				
vive; lavoura	Lavoura	Agricultura	x				x
vive; lavoura; gados; criados	Lavoura	Agricultura		x			
vive; letras	Artes liberais	Justiça		x			
vive; letras; fazendas	Artes liberais	Justiça		x			
vive; letras; medicina	Artes liberais	Saúde		x			
vive; lucros; curativo	Artes liberais	Saúde		x			
vive; lucros; mercearia	Mercancia	Negócio	x				
vive; missa	Igreja	Eclesiástico		x			
vive; negócio	Mercancia	Negócio		x			
Vive; negócio; fazenda; vinhos	Mercancia	Negócio	x				
vive; negócio; fazendas; miúdas	Mercancia	Negócio	x				
vive; negócio; solicitador; causas	Mercancia	Negócio	x				
vive; ordens; bens; patrimoniais	Igreja	Eclesiástico		x			
vive; pano; linho	Artes mecânicas	Transformação	x				
vive; propriedades	Lavoura	Agricultura		x			
vive; quotidiano; trabalho	Lavoura	Agricultura	x				
vive; renda	Viver da sua fazenda	Viver da sua fazenda		x			
vive; rendas	Viver da sua fazenda	Viver da sua fazenda		x			
vive; rendas; morgado	Viver da sua fazenda	Viver da sua fazenda		x			
vive; rendas; património	Viver da sua fazenda	Viver da sua fazenda		x			
vive; rendas; quintas	Lavoura	Agricultura		x			
vive; rendas; toma	Mercancia	Negócio		x			
vive; rendimento; letras	Artes liberais	Viver da sua fazenda		x			x
vive; rendimentos	Viver da sua fazenda	Viver da sua fazenda		x			
vive; searas	Lavoura	Agricultura	x				
vive; soldos	Milícia	Defesa		x			
vive; suas; terras	Lavoura	Agricultura		x			
vive; trabalho	Lavoura	Agricultura	x				
vive; trabalho; escritório	Governo da res publica	Administração		x			
vive; vários; contratos	Mercancia	Negócio		x			
vive; vários; negócios	Mercancia	Negócio		x			
vivia; loja	Mercancia	Negócio	x				
volanteiro	Artes mecânicas	Transformação	x				

Me-Mecânico; NM-Não Mecânico; No-Nobre; NS-Não se sabe



# C

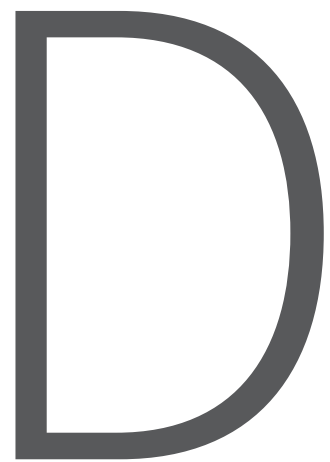
Lista de *stopwords*

a	através	dar	distrito	esteve
à	azurar	das	dito	estive
abaixo	bairro	de	diz	estivemos
acerca	barroso	debaixo	dizem	estiver
acima	baixo	decerto	dizer	estivera
adeus	bastante	defronte	dois	estiveram
ágil	beira	dela	donde	estivéramos
agora	bem	delas	dos	estiverem
aí	boa	dele	douro	estivermos
ainda	boas	deles	doze	estivesse
além	bom	demais	duas	estivessem
algo	bons	dentro	durante	estivéssemos
alguém	braga	depois	dúvida	estiveste
alguma	breve	depressa	e	estivestes
algumas	cá	desde	é	estou
alguns	cada	desligado	ela	estrangeiras
algures	caminho	dessa	elas	eu
ali	canas	dessas	ele	exceto
amanhã	caniçada	desse	eles	exemplo
ambos	catorze	desses	embaixo	faço
amiúde	cedo	desta	embora	falta
ampla	cento	destas	enfim	fará
amplas	certamente	deste	enquanto	favor
amplo	certeza	destes	então	fazeis
amplos	cima	detrás	entre	fazem
ano	cinco	devagar	era	fazemos
anos	coisa	deve	eram	fazendo
ante	coisas	devem	éramos	fazes
antes	com	devendo	és	feita
aonde	como	dever	essa	feitas
aos	comprido	deverá	essas	feito
apenas	conhecido	deverão	esse	felgueiras
apoio	contra	deveras	esses	fez
apontar	contudo	deveria	está	fim
após	corrente	deveriam	esta	final
aquela	covas	devia	estamos	foi
aquelas	cuja	deviam	estão	fomos
aquele	cujas	dez	estar	for
aqueles	cujo	dezanove	estas	fora
aqui	cujos	dezasseis	estás	foram
aquilo	custa	dezassete	estava	fôramos
arcos	dá	dezoito	estavam	forem
área	dão	dia	estávamos	forma
as	daquela	diante	este	formos
às	daquelas	direita	esteja	fosse
assim	daquele	disse	estejam	fossem
até	daqueles	disso	estejamos	fôssemos
atrás	daquilo	disto	estes	foste

fostes	jamais	nela	oito	porque
foz	la	nelas	onde	porquê
freguesia	lá	nele	ontem	portanto
freixada	lado	neles	onze	porto
fui	lamego	nem	os	posição
funchal	lhe	nenhum	ou	possível
gondomar	lhes	nenhuma	outra	possivelmente
grandes	lima	nenhumas	outras	posso
grupo	lo	nenhuns	outrem	pouca
há	local	nenhures	outro	poucas
haja	logo	nessa	outrora	pouco
hajam	longe	nessas	outros	poucos
hajamos	lugar	nesse	parece	primeira
hão	maçal	nesses	paroquial	primeiras
havemos	maioria	nesta	parte	primeiro
havia	maiorias	nestas	partir	primeiros
hei	mais	neste	paucas	primeiro
hoje	mal	nestes	pegar	própria
hora	mas	ninguém	pela	próprias
horas	máximo	nisso	pelas	próprio
houve	me	nisto	pelo	próprios
houvemos	menos	nível	pelos	próxima
houver	mês	no	penitenciária	próximas
houvera	meses	noite	pequena	próximo
houverá	mesma	nome	pequenas	próximos
houveram	mesmas	nos	pequeno	pude
houvéramos	mesmo	nós	pequenos	puderam
houverão	mesmos	nossa	per	quais
houverei	meu	nossas	perante	qual
houverem	meus	nosso	perto	qualquer
houveremos	mil	nossos	peessoas	quando
houveria	minha	nova	pode	quanto
houveriam	minhas	novas	pôde	quantos
houveríamos	momento	nove	podem	quarta
houvermos	moncorvo	novo	podendo	quarto
houvesse	muita	novos	poder	quatro
houvessem	muitas	num	poderá	que
houvéssemos	muito	numa	poderia	quê
iniciar	muitos	numas	poderiam	quem
inicio	na	número	podia	quer
ir	nada	nunca	podiam	quereis
irá	não	nuns	põe	querem
irmandade	naquela	o	põem	queremas
isso	naquelas	obra	pois	queres
ista	naquele	obrigada	ponto	quero
iste	naqueles	obrigado	pontos	questão
isto	naquilo	oitava	por	quieto
já	nas	oitavo	porém	quinta

quintela	sois	tido	vêm
quinto	somente	tinha	vendo
quinze	somos	tinham	vens
real	sou	tínhamos	ver
relação	soutelo	tipo	verdade
resende	sua	tive	verdadeiro
ribeira	suas	tivemos	vez
s.	tal	tiver	vezes
sabe	talvez	tivera	viagem
sabem	também	tiveram	viana
saber	tampouco	tivéramos	vindo
sanfins	tanta	tiverem	vinte
santa	tantas	tivermos	vir
são	tanto	tivesse	viseu
são	tantos	tivessem	você
se	tão	tivéssemos	vocês
segunda	tarde	tiveste	vos
segundo	te	tivestes	vós
sei	tem	toda	vossa
seisseja	tém	todas	vossas
sejam	têm	todavia	vosso
sejamos	temos	todo	vossos
sem	tempo	todos	zero
sempre	tendes	trás	
sendo	tendo	três	
senhorim	tenha	treze	
ser	tenham	tu	
será	tenhamos	tua	
serão	tenho	tuas	
serei	tens	tudo	
seremos	tentar	última	
seria	tentaram	últimas	
seriam	tente	último	
seríamos	tentei	últimos	
sete	ter	um	
sétima	terá	uma	
sétimo	terão	umas	
seu	terceira	unhão	
seus	terceiro	uns	
sexta	terei	usar	
sexto	teremos	vai	
si	teria	vais	
sido	teriam	valdevez	
sim	teríamos	valor	
sistema	teu	vandoma	
só	teus	vão	
sob	teve	veja	
sobre	ti	vem	





Eventos sem *stopwords*, com  
aplicação de *stemming*

Evento	tsVector
Vive de algumas fazendas e granjarias de gados e lavouras	'fazend':4 'gad':8 'granj':6 'lavour':10 'viv':1
Vive de algumas fazendas que tem na sua pátria	'fazend':4 'pátr':9 'viv':1
Vive de alugar panos de seda	'alug':3 'pan':4 'sed':6 'viv':1
Vive de ensinar meninos	'ensin':3 'menin':4 'viv':1
Vive de fazendas que possui de raíz	'fazend':3 'possu':5 'raíz':7 'viv':1
Vive de fazer panos	'faz':3 'pan':4 'viv':1
Vive de guardar gado	'gad':4 'guard':3 'viv':1
Vive de pano de linho	'linh':5 'pan':3 'viv':1
Vive de rendas da casa de seus pais	'cas':5 'da':4 'pais':8 'rend':3 'viv':1
Vive de rendas que toma	'rend':3 'tom':5 'viv':1
Vive de seu contrato	'contrat':4 'viv':1
Vive de seu contrato de venda de chapéus, vinhos e aguardente	'aguardent':11 'chapéus':8 'contrat':4 'vend':6 'vinh':9 'viv':1
Vive de seu grangeio	'grangei':4 'viv':1
Vive de seu granjeio	'granjei':4 'viv':1
Vive de seu jornal e trabalho da sua roca	'da':7 'jornal':4 'roc':9 'trabalh':6 'viv':1
Vive de seu negócio	'negóci':4 'viv':1
Vive de seu ofício de conteiro	'conteir':6 'ofíci':4 'viv':1
Vive de seu trabalho	'trabalh':4 'viv':1
Vive de seu trabalho a agência	'agênc':6 'trabalh':4 'viv':1
Vive de seu trabalho e agência	'agênc':6 'trabalh':4 'viv':1
Vive de seu trabalho e agência e de fazer algumas searas	'agênc':6 'faz':9 'sear':11 'trabalh':4 'viv':1
Vive de seu trabalho e fazenda	'fazend':6 'trabalh':4 'viv':1
Vive de seus bens	'bens':4 'viv':1
Vive de seus jornais	'jorn':4 'viv':1
Vive de seus morgadios	'morgadi':4 'viv':1
Vive de sua agência	'agênc':4 'viv':1
Vive de sua agência e contrato	'agênc':4 'contrat':6 'viv':1
Vive de sua agência e contrato de sola e gados e dos ordenados que tem de procurador	'agênc':4 'contrat':6 'gad':10 'orden':13 'procur':17 'sol':8 'viv':1
Vive de sua agência e fazenda	'agênc':4 'fazend':6 'viv':1
Vive de sua agência e trabalho	'agênc':4 'trabalh':6 'viv':1
Vive de sua fazenda em Machico	'em':5 'fazend':4 'machic':6 'viv':1
Vive de sua fazenda fidalgamente	'fazend':4 'fidalg':5 'viv':1
Vive de sua fazenda, sem ofício algum	'algum':7 'fazend':4 'ofíci':6 'viv':1
Vive de sua industria	'industr':4 'viv':1
Vive de sua indústria	'indústr':4 'viv':1
Vive de sua indústria e trabalho	'indústr':4 'trabalh':6 'viv':1
Vive de sua lavoura	'lavour':4 'viv':1
Vive de sua lavoura com seus gados e criados	'cri':9 'gad':7 'lavour':4 'viv':1
Vive de sua loja	'loj':4 'viv':1
Vive de sua loja de mercearia	'loj':4 'merc':6 'viv':1
Vive de suas agências	'agênc':4 'viv':1
Vive de suas curiosidades	'curios':4 'viv':1
Vive com a fazenda de seus pais	'fazend':4 'pais':7 'viv':1
Vive com o trato das suas lavouras	'lavour':7 'trat':4 'viv':1
Vive com o trato de carreteiro	'carreteir':6 'trat':4 'viv':1
Vive com o trato de estaqueiro	'estaqueir':6 'trat':4 'viv':1

Evento	tsVector
Vive com o trato de sua agência	'agênc':7 'trat':4 'viv':1
Vive com o trato de suas lavouras	'lavour':7 'trat':4 'viv':1
Vive com sua agência	'agênc':4 'viv':1
Vive da agência de algum negócio	'agênc':3 'algum':5 'da':2 'negóci':6 'viv':1
Vive da sua agência	'agênc':4 'da':2 'viv':1
Vive da sua agência e de contratar em ferro de que tem loja	'agênc':4 'contrat':7 'da':2 'em':8 'ferr':9 'loj':13 'viv':1
Vive da sua agência e negociação de vinhos que manda vender por sua conta	'agênc':4 'cont':14 'da':2 'mand':10 'negoc':6 'vend':11 'vinh':8 'viv':1
Vive da sua fazenda	'da':2 'fazend':4 'viv':1
Vive da sua fazenda e agência	'agênc':6 'da':2 'fazend':4 'viv':1
Vive da sua fazenda e indústria	'da':2 'fazend':4 'indústr':6 'viv':1
Vive da sua fazenda e morgado	'da':2 'fazend':4 'morg':6 'viv':1
Vive da sua fazenda e negócio de mercância	'da':2 'fazend':4 'mercânc':8 'negóci':6 'viv':1
Vive da sua fazenda na realidade é feitor do Conde do Sabugal	'cond':10 'da':2 'do':9,11 'fazend':4 'feitor':8 'realidad':6 'sabugal':12 'viv':1
Vive da sua indústria	'da':2 'indústr':4 'viv':1
Vive da sua lavoura	'da':2 'lavour':4 'viv':1
Vive do trato de fazer imagens de santos	'do':2 'faz':5 'imagens':6 'sant':8 'trat':3 'viv':1
Vive dos lucros do seu curativo	'curat':6 'do':4 'lucr':3 'viv':1
Vive dos lucros que tira da sua mercearia	'da':6 'lucr':3 'merc':8 'tir':5 'viv':1
Vive dos rendimentos das suas fazendas	'fazend':6 'rendiment':3 'viv':1
Vive dos rendimentos de suas fazendas que são muitas e boas	'fazend':6 'rendiment':3 'viv':1
Vive dos rendimentos do seu morgado	'do':4 'morg':6 'rendiment':3 'viv':1
Vive dos rendimentos dos seus morgados	'morg':6 'rendiment':3 'viv':1
Vive dos seus bens	'bens':4 'viv':1
Vive da sua loja de mercador que é grossa e de lãs que fazia	'da':2 'faz':14 'gross':9 'loj':4 'lãs':12 'mercador':6 'viv':1
Vive da sua missa	'da':2 'miss':4 'viv':1
Vive das duas fazendas	'fazend':4 'viv':1
Vive das fazendas que possui nos subúrbios da cidade de Coimbra	'cidad':9 'coimbr':11 'da':8 'fazend':3 'possu':5 'subúrb':7 'viv':1
Vive das rendas do seu morgado	'do':4 'morg':6 'rend':3 'viv':1
Vive das suas fazendas	'fazend':4 'viv':1
Vive das suas fazendas e lavoura com grosso trato	'fazend':4 'gross':8 'lavour':6 'trat':9 'viv':1
Vive das suas fazendas e lavouras	'fazend':4 'lavour':6 'viv':1
Vive das suas fazendas e morgadio	'fazend':4 'morgadi':6 'viv':1
Vive das suas fazendas e morgados	'fazend':4 'morg':6 'viv':1
Vive das suas fazendas e rendas	'fazend':4 'rend':6 'viv':1
Vive das suas fazendas fabricando-as por seus criados e jornaleiros	'cri':10 'fabric':6 'fabricando':5 'fazend':4 'jornaleir':12 'viv':1
Vive das suas fazendas gados e lavouras	'fazend':4 'gad':5 'lavour':7 'viv':1
Vive das suas fazendas que cultiva	'cultiv':6 'fazend':4 'viv':1
Vive das suas fazendas que são vastas	'fazend':4 'vast':7 'viv':1
Vive das suas lavouras	'lavour':4 'viv':1
Vive das suas letras	'letr':4 'viv':1
Vive das suas rendas	'rend':4 'viv':1
Vive das suas terras	'terr':4 'viv':1
Vive de algum contrato	'algum':3 'contrat':4 'viv':1

Evento	tsVector
Vive de sua fazenda	'fazend':4 'viv':1
Vive de sua fazenda - foi sapateiro	'fazend':4 'sapateir':6 'viv':1
Vive de sua fazenda e agência	'agênc':6 'fazend':4 'viv':1
Vive de sua fazenda e contratos	'contrat':6 'fazend':4 'viv':1
Vive de sua fazenda e de sua botica	'botic':8 'fazend':4 'viv':1
Vive de sua fazenda e rendas	'fazend':4 'rend':6 'viv':1
Vive de sua fazenda e teve o ofício de pedreiro	'fazend':4 'ofíci':8 'pedreir':10 'viv':1
Vive de sua fazenda e trabalho	'fazend':4 'trabalh':6 'viv':1
Vive de sua fazenda e trata em rendas	'em':7 'fazend':4 'rend':8 'trat':6 'viv':1
Vive dos seus bens e fazendas	'bens':4 'fazend':6 'viv':1
Vive dos seus contratos de renda	'contrat':4 'rend':6 'viv':1
Vive dos seus fiéis	'fié':4 'viv':1
Vive dos seus morgados	'morg':4 'viv':1
Vive dos seus rendimentos	'rendiment':4 'viv':1
Vive dos seus soldos	'sold':4 'viv':1
Vive limpa e abastadamente de sua fazenda	'abast':4 'fazend':7 'limp':2 'viv':1
Vive limpamente de seu negócio	'limp':2 'negóci':5 'viv':1
Vive nesta data sem ocupação alguma	'dat':3 'ocup':5 'viv':1
Vive por seus bens e fazendas	'bens':4 'fazend':6 'viv':1
Vive por sua fazenda	'fazend':4 'viv':1
Vive por sua fazenda e exercício no ofício de esteireiro	'esteireir':10 'exercíci':6 'fazend':4 'ofíci':8 'viv':1
Vive se sua lavoura e fazenda	'fazend':6 'lavour':4 'viv':1
Vive sem ofício	'ofíci':3 'viv':1
Vive também de sua fazenda	'fazend':5 'viv':1
Vive de suas fazendas	'fazend':4 'viv':1
Vive de suas fazendas como lavrador	'fazend':4 'lavrador':6 'viv':1
Vive de suas fazendas e governança	'fazend':4 'governanc':6 'viv':1
Vive de suas fazendas e lavoura	'fazend':4 'lavour':6 'viv':1
Vive de suas fazendas e morgadios	'fazend':4 'morgadi':6 'viv':1
Vive de suas fazendas e morgado	'fazend':4 'morg':6 'viv':1
Vive de suas fazendas e negócios	'fazend':4 'negóci':6 'viv':1
Vive de suas fazendas e rendas	'fazend':4 'rend':6 'viv':1
Vive de suas fazendas e rendimentos e dinheiros à razão de juros	'dinheir':8 'fazend':4 'jur':12 'razã':10 'rendiment':6 'viv':1 'à':9
Vive de suas fazendas limpa e abastadamente	'abast':7 'fazend':4 'limp':5 'viv':1
Vive de suas fazendas sem ter ofício	'fazend':4 'ofíci':7 'viv':1
Vive de suas lavouras	'lavour':4 'viv':1
Vive de suas lavouras e fazenda	'fazend':6 'lavour':4 'viv':1
Vive de suas letras sendo consultado por várias pessoas	'consult':6 'letr':4 'viv':1 'vár':8
Vive de suas searas	'sear':4 'viv':1
Vive de suas terras	'terr':4 'viv':1
Vive de tomar uma renda do Bispal da vila de Sarzedas	'bispal':7 'da':8 'do':6 'rend':5 'sarzed':11 'tom':3 'vil':9 'viv':1
Vive de uma capela	'capel':4 'viv':1
Vive de uma lavoura	'lavour':4 'viv':1
Vive de vários contratos	'contrat':4 'viv':1 'vári':3
Vive de vários negócios além da sua ocupação de sacristão	'da':6 'negóci':4 'ocup':8 'sacristã':10 'viv':1 'vári':3
Vive do contracto de panos de cor	'contract':3 'cor':7 'do':2 'pan':5 'viv':1

Evento	tsVector
Vive do contrato de comprar e vender cera	'cer':8 'compr':5 'contrat':3 'do':2 'vend':7 'viv':1
Vive do contrato de rendas	'contrat':3 'do':2 'rend':5 'viv':1
Vive do ofício de alfaiate de que é mestre e tem sua loja de pano de linho	'alfaiat':5 'do':2 'linh':17 'loj':13 'mestr':9 'ofíci':3 'pan':15 'viv':1
Vive do rendimento das suas letras	'do':2 'letr':6 'rendiment':3 'viv':1
Vive do ofício de levar cartas do correio de Arraiolos ao correio de Évora	'ao':11 'arraiol':10 'cart':6 'correi':8,12 'do':2,7 'lev':5 'ofíci':3 'viv':1 'évor':14
Vive do rendimento das suas fazendas	'do':2 'fazend':6 'rendiment':3 'viv':1
Vive do rendimento das suas fazendas que constam de vinhas	'const':8 'do':2 'fazend':6 'rendiment':3 'vinh':10 'viv':1
Vive do rendimento das suas Letras e da Medicina	'da':8 'do':2 'letr':6 'medicin':9 'rendiment':3 'viv':1
Vive do rendimento das suas letras e fazendas	'do':2 'fazend':8 'letr':6 'rendiment':3 'viv':1
Vive do rendimento de fazendas que tem comprado	'compr':8 'do':2 'fazend':5 'rendiment':3 'viv':1
Vive do seu contrato de marceria	'contrat':4 'do':2 'marc':6 'viv':1
Vive do seu contrato e agência	'agênc':6 'contrat':4 'do':2 'viv':1
Vive do seu granjeamento	'do':2 'granjeament':4 'viv':1
Vive do seu manejo	'do':2 'manei':4 'viv':1
Vive do seu morgado e fazendas	'do':2 'fazend':6 'morg':4 'viv':1
Vive do seu negócio	'do':2 'negóci':4 'viv':1
Vive do seu negócio com loja de várias fazendas miúdas	'do':2 'fazend':9 'loj':6 'miúd':10 'negóci':4 'viv':1 'vár':8
Vive do seu negócio e de algum bocado de fazenda que tem e de ser provador de vinhos dos estrangeiros	'algum':7 'boc':8 'do':2 'estrangeir':20 'fazend':10 'negóci':4 'provador':16 'vinh':18 'viv':1
Vive do seu quotidiano trabalho	'do':2 'quotidian':4 'trabalh':5 'viv':1
Vive do seu trabalho	'do':2 'trabalh':4 'viv':1
Vive do seu trabalho e agência	'agênc':6 'do':2 'trabalh':4 'viv':1
Vive do trabalho da sua almofada	'almof':6 'da':4 'do':2 'trabalh':3 'viv':1
Vive do trabalho no seu escritório	'do':2 'escritóri':6 'trabalh':3 'viv':1
Vive do trato de mercador	'do':2 'mercador':5 'trat':3 'viv':1



E

Protótipo em desenvolvimento

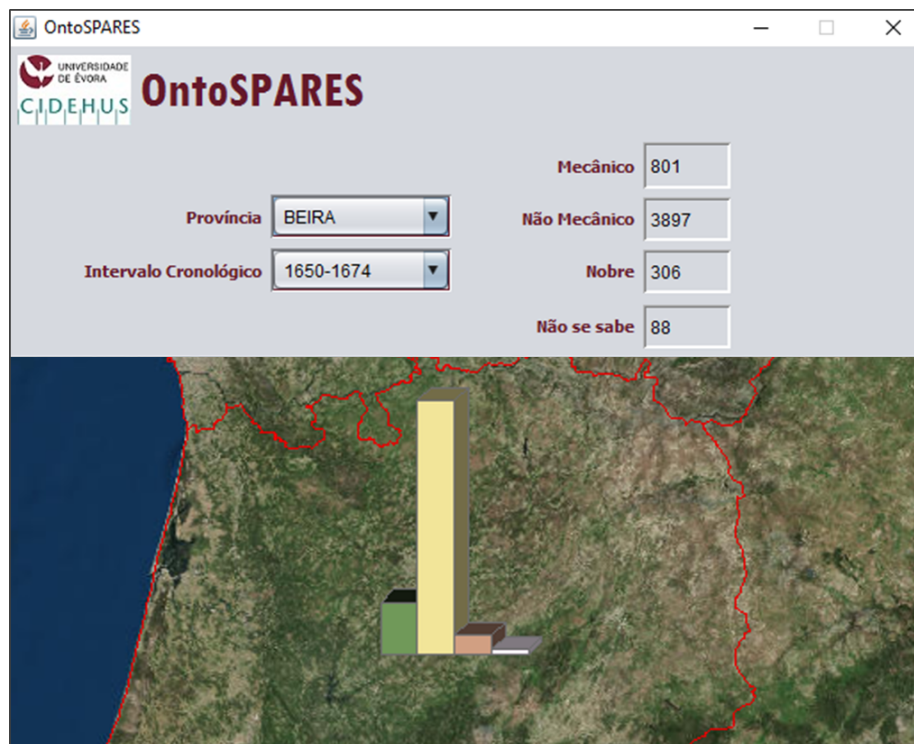


Figura E.1: Distribuição dos estatutos para a província da Beira.

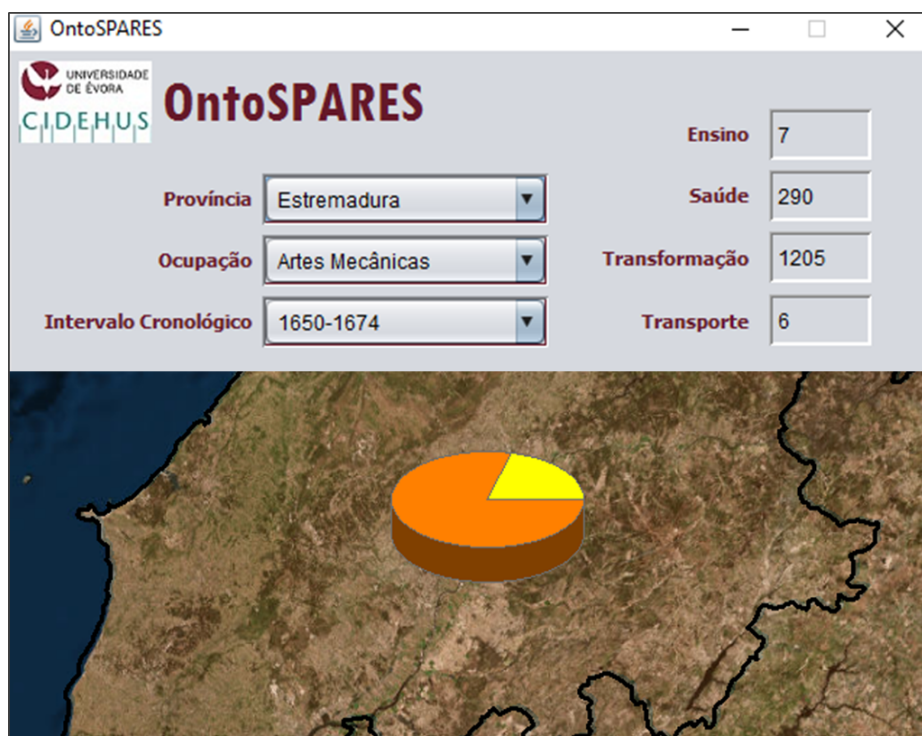


Figura E.2: Contabilização dos setores de atividade para a província da Estremadura e ocupação Artes Mecânicas.





Figura E.3: Distribuição, por província, do ofício “Vive de sua fazenda”.



# Referências bibliográficas

- Agustini, A. (2006). *Aquisição Automática de Subcategorização Sintático-Semântica e sua Utilização em Sistemas de Processamento da Língua Natural*. PhD thesis, Universidade Nova de Lisboa, Faculdade de Ciências e Tecnologia. Disponível em [http://dev-htl.di.fct.unl.pt/gpl/projects/PATRAS/tese\\_agustini\\_10mai.pdf](http://dev-htl.di.fct.unl.pt/gpl/projects/PATRAS/tese_agustini_10mai.pdf) [Consultado em 2014-12-19].
- Anoop, K. & Lajish, V. L. (2015). Morphology based text detection and extraction from Malayalam news videos. *National Conference on Indian Language Computing (NCILC - 2015)*, pp. 1–6.
- Antoniou, G. & Van Harmelen, F. (2004). Web ontology language: Owl. In Staab, S. & Studer, R., editors, *Handbook on ontologies*, pp. 67–92. Springer Berlin Heidelberg. doi:10.1007/978-3-540-24750-0\_4.
- Antoniou, G. & Van Harmelen, F. (2008). *A Semantic Web Primer*. MIT Press, Cambridge MA, USA, 2 edition.
- Artmann, S. (2013). Pragmatism and the Evolution of Semantic Systems. In *Evolution of Semantic Systems*, pp. 13–29. Springer. doi:10.1007/978-3-642-34997-3.
- Aundhakar, S. P. & Pokale, N. B. (2015). A Survey on User Personalization using Ontology. *International Journal of Science and Research*, 4(1):696–699. Disponível em <http://www.ijsr.net/archive/v4i1/SUB15222.pdf> [Consultado em 2015-06-18].
- Bajenaru, L. & Smeureanu, I. (2015). An Ontology Based Approach for Modeling E-Learning in Healthcare Human Resource Management. *Economic Computation & Economic Cybernetics Studies & Research*, 49(1):18–34. Disponível em [http://www.ecocyb.ase.ro/nr20151/02-BajenaruLidia,IonSmeureanu\(T\).pdf](http://www.ecocyb.ase.ro/nr20151/02-BajenaruLidia,IonSmeureanu(T).pdf) [Consultado em 2015-07-05].
- Bao, Y., Tian, Q., Chen, M., & Lin, H. (2015). An automatic extraction method for individual tree crowns based on self-adaptive mutual information and tile computing. *International Journal of Digital Earth*, 8(6):495–516. doi:10.1080/17538947.2014.912683.
- Baoxian, J. & Yang, W. (2013). Dynamic Forensics Model based on Ontology and Context Information. *International Journal of Computer Science Issues*, 10(2):270–272. Disponível em <http://ijcsi.org/papers/IJCSI-10-2-2-270-272.pdf> [Consultado em 2014-12-19].
- Baranow, U. G. (1983). Perspectivas na Contribuição da Linguística e de Áreas afins à Ciência da Informação. *Ciência Da Informação*, 12(1):23–35. Disponível em <http://revista.ibict.br/ciinf/article/view/191/191> [Consultado em 2014-12-19].

- Bird, S., Klein, E., & Loper, E. (2008). Introduction to Language Processing and Python. Disponível em <http://nltk.sourceforge.net/doc/en/ch01.html> [Consultado em 2014-12-19].
- Botelho, T. R., van Leeuwen, M. H. D., Maas, I., & Miles, A. (2006). HISCO (Historical International Standard Classification of Occupation): construindo uma codificacao de ocupacoes para o passado brasileiro. *Revista da ABET*, 6(2):166–181. Disponível em <http://periodicos.ufpb.br/ojs/index.php/abet/article/view/15722/8967> [Consultado em 2016-04-28].
- Caldeira, C. P. (2011). *A Arte das Bases de Dados*. Edições Sílabo, Lisboa, 1 edition.
- Calvanese, D., Cogrel, B., Komla-Ebri, S., Kontchakov, R., Lanti, D., Rezk, M., Rodriguez-Muro, M., & Xiao, G. (2016). Ontop: Answering SPARQL queries over relational databases. *Semantic Web*, (Preprint):1–17. doi:10.3233/SW-160217.
- Chaudhari, P. & Vankudothu, B. (2015). Flexible and improved method for automatic Semantic Content extraction in videos. In *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, pp. 1–4, Coimbatore. IEEE. doi:10.1109/ICIIECS.2015.7193113.
- Chen, H., Han, D., Dai, Y., & Zhao, L. (2015). Design of automatic extraction algorithm of knowledge points for MOOCs. *Computational Intelligence and Neuroscience*, 2015:1–10. doi:10.1155/2015/123028.
- Chisholm, A. (2013). *Exploring Data with RapidMiner*. Packt Publishing Ltd.
- Chopra, A., Prashar, A., & Sain, C. (2013). Natural language processing. *International Journal of Technology Enhancements and Emerging Engineering Research*, 1(4):131–134.
- Choudhury, S. R. & Giles, C. L. (2015). An Architecture for Information Extraction from Figures in Digital Libraries. In *Proceedings of the 24th International Conference on World Wide Web, WWW '15 Companion*, pp. 667–672, Florence. ACM. doi:10.1145/2740908.2741712.
- Conde, A., Larrañaga, M., Arruarte, A., Elorriaga, J. A., & Roth, D. (2016). litewi: A combined term extraction and entity linking method for eliciting educational ontologies from textbooks. *Journal of the Association for Information Science and Technology*, 67(2):380–399. doi:10.1002/asi.23398.
- Dale, R. (2010). Classical Approaches to Natural Language Processing. In Indurkha, N. & Damerau, F., editors, *Handbook of Natural Language Processing*, chapter 1, pp. 3–8. CRC Press, Boca Raton, 2 edition.
- Das, N. N. & Kumar, E. (2014). Identification of Query Forms for Retrieving the Information From Deep Web. *Transactions on Machine Learning and Artificial Intelligence*, 2(6):53–61. doi:10.14738/tmlai.26.778.
- Decker, S., Melnik, S., Van Harmelen, F., Fensel, D., Klein, M., Broekstra, J., Erdmann, M., & Horrocks, I. (2000). The semantic web: The roles of XML and RDF. *IEEE Internet computing*, 4(5):63–73. doi:10.1109/4236.877487.
- Dhingra, V. & Bhatia, K. K. (2015). Development of Ontology in Laptop Domain for Knowledge Representation. In *Procedia Computer Science*, volume 46, pp. 249–256. doi:10.1016/j.procs.2015.02.018.
- Di Felippo, A. & Dias-da Silva, B. C. (2009). O Processamento Automático de Línguas Naturais enquanto Engenharia do Conhecimento Linguístico. *Calidoscópio*, 7(3):183–191. doi:10.4013/cld.2009.73.02.
- Dieb, T. M., Yoshioka, M., Hara, S., & Newton, M. C. (2015). Framework for automatic information extraction from research papers on nanocrystal devices. *Beilstein Journal of Nanotechnology*, 6:1872–1882. doi:10.3762/bjnano.6.190.
- Efremova, J., García, A. M., Zhang, J., & Calders, T. (2015). Towards population reconstruction: extraction of family relationships from historical documents. In *First International Workshop on Population Informatics for Big Data (21th ACM-SIGKDD PopInfo'15)*, pp. 1–9, Sydney.

- Elsayed, A.-E., El-Beltagy, S. R., Rafea, M., & Hegazy, O. (2007). Applying Data Mining for Ontology Building. In *IEEE International Workshop on Security in e-Science and e-Research (ISSR)*, pp. 1–14. Disponível em [http://www.arc.sci.eg/NARIMS\\_upload/NARIMSDocs/73180/OntologyBuilding.pdf](http://www.arc.sci.eg/NARIMS_upload/NARIMSDocs/73180/OntologyBuilding.pdf) [Consultado em 2013-05-27].
- Ford, E., Carroll, J. A., Smith, H. E., Scott, D., & Cassell, J. A. (2016). Extracting information from the text of electronic medical records to improve case detection: a systematic review. *Journal of the American Medical Informatics Association*. doi:10.1093/jamia/ocv180.
- Fragkou, P. (2013). Information extraction versus text segmentation for web content mining. *International Journal of Software Engineering and Knowledge Engineering*, 23(08):1109–1137. doi:10.1142/S0218194013500332.
- Fraihat, S. & Shambour, Q. (2015). A Framework of Semantic Recommender System for e-Learning. *Journal of Software*, 10(3):317–330. doi:10.17706/jsw.10.3.317-330.
- Gaizauskas, R., Saggion, H., & Barker, E. (2007). Information Access and Natural Language Processing: a Stimulating Dialogue. In Ahmad, K., Brewster, C., & Stevenson, M., editors, *Words and Intelligence II – Essays in Honor of Yorick Wilks*, pp. 85–105. Springer, Netherlands. doi:10.1007/1-4020-5833-0\_4.
- Gašević, D., Djuric, D., & Devedžić, V. (2006). *Model Driven Architecture and Ontology Development*. Springer-Verlag Berlin Heidelberg. doi:10.1007/3-540-32182-9.
- Goddard, C. & Schalley, A. (2010). Semantic Analysis. In Indurkha, N. & Damerau, F., editors, *Handbook of Natural Language Processing*, chapter 5, pp. 93–120. CRC Press, Boca Raton, 2 edition.
- Grishman, R. (1986). *Computational Linguistics: An Introduction*. Studies in Natural Language Processing. Cambridge University Press.
- Gruber, T. (2009). Ontology. In Liu, L. & Özsu, M., editors, *Encyclopedia of Database Systems SE - 1318*, pp. 1963–1965. Springer US. doi:10.1007/978-0-387-39940-9\_1318.
- Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2):199–220. doi:10.1.1.101.7493.
- Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43(5-6):907–928. doi:10.1006/ijhc.1995.1081.
- Guarino, N. (1998). *Formal ontology in information systems*. IOS Press Amsterdam, The Netherlands. Disponível em <http://osm.cs.byu.edu/CS652s04/Gua98Formal.pdf> [Consultado em 2014-12-14].
- Guarino, N., Oberle, D., & Staab, S. (2009). *What is an Ontology?* Springer Berlin Heidelberg. doi:10.1007/978-3-540-92673-3\_0.
- Halioui, A., Valtchev, P., & Diallo, A. B. (2015). Acquisition of Generic Problem Solving Knowledge through Information Extraction and Pattern Mining. In *2015 IEEE 27th International Conference on Tools with Artificial Intelligence*, pp. 583–590, Vietri sul Mare. doi:10.1109/ICTAI.2015.90.
- Han, D., Shi, Y., Wang, W., & Dai, Y. (2013). Research on Multi-level Association Rules Based on Geosciences Data. *Journal of Software*, 8(12):3269–3276. doi:10.4304/jsw.8.12.3269-3276.
- Hanks, P. (2013). *Lexical analysis: Norms and exploitations*. Mit Press. doi:10.1093/ijl/ect025.
- Hartig, O. & Heese, R. (2007). The SPARQL query graph model for query optimization. In *European Semantic Web Conference*, pp. 564–578. Springer, Springer Berlin Heidelberg. doi:10.1007/978-3-540-72667-8\_40.

- Hazber, M. A. G., Li, R., Gu, X., Xu, G., & Li, Y. (2015). Semantic SPARQL query in a relational database based on ontology construction. In *2015 11th International Conference on Semantics, Knowledge and Grids (SKG)*, pp. 25–32, Beijing, China. IEEE. doi:10.1109/SKG.2015.14.
- Hippisley, A. (2010). Lexical Analysis. In Indurkha, N. & Damerau, F., editors, *Handbook of Natural Language Processing*, chapter 3, pp. 31–58. CRC Press, Boca Raton, 2 edition.
- Horch, A., Kett, H., & Weisbecker, A. (2015). Extracting product unit attributes from product offers by using an ontology. In *2015 Second International Conference on Computer Science, Computer Engineering, and Social Media (CSCESM)*, pp. 67–71, Lodz. IEEE. doi:10.1109/CSCESM.2015.7331830.
- Horridge, M. (2011). *A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools Edition 1.3*. The University Of Manchester. Disponível em [https://mariajulianadascalu.files.wordpress.com/2014/02/owl-cs-manchester-ac-uk\\_-eowltutorialp4\\_v1\\_3.pdf](https://mariajulianadascalu.files.wordpress.com/2014/02/owl-cs-manchester-ac-uk_-eowltutorialp4_v1_3.pdf) [Consultado em 2015-04-17].
- Ide, N. & Woolner, D. (2007). Historical ontologies. In Ahmad, K., Brewster, C., & Stevenson, M., editors, *Words and Intelligence II*, volume 36, pp. 137–152. Springer Netherlands. doi:10.1007/1-4020-5833-0\_7.
- Javed, M., Nagabhushan, P., & Chaudhuri, B. B. (2015). Automatic extraction of correlation-entropy features for text document analysis directly in run-length compressed domain. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1–5, Tunis. IEEE. doi:10.1109/ICDAR.2015.7333714.
- Knublauch, H., Fergerson, R. W., Noy, N. F., & Musen, M. A. (2004). The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications. In McIlraith, S. A., Plexousakis, D., & van Harmelen, F., editors, *The Semantic Web – ISWC 2004*, volume 3298 of *Lecture Notes in Computer Science*, pp. 229–243. Springer Berlin Heidelberg. doi:10.1007/978-3-540-30475-3\_17.
- Kollia, I., Glimm, B., & Horrocks, I. (2011). *SPARQL Query Answering over OWL Ontologies*, pp. 382–396. Springer Berlin Heidelberg, Berlin, Heidelberg. doi:10.1007/978-3-642-21034-1\_26.
- Kreuzthaler, M., Schulz, S., & Berghold, A. (2015). Secondary use of electronic health records for building cohort studies through top-down information extraction. *Journal of Biomedical Informatics*, 53:188–195. doi:10.1016/j.jbi.2014.10.010.
- Kumar, J., Reddy, S., Priya, M., Raju, J., & Vanmathi, C. (2013). Implementation Of Ontology Matching Using Protégé. *International Journal of Computer Applications Technology and Research*, 2(6):723–725. doi:10.7753/IJCATR0206.1017.
- Küppers, B.-O. (2013). Elements of a semantic code. In Küppers, B.-O., Hahn, U., & Artmann, S., editors, *Evolution of Semantic Systems*, pp. 67–85. Springer, 1 edition. doi:10.1007/978-3-642-34997-3\_4.
- Lahbib, W., Bounhas, I., & Slimani, Y. (2015). Arabic Terminology Extraction and Enrichment Based on Domain-Specific Text Mining. In *2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 340–347, Vietri sul Mare. IEEE. doi:10.1109/ICTAI.2015.59.
- Liddy, E. D. (2001). Natural language processing. *Encyclopedia of Library and Information Science*. Disponível em <http://surface.syr.edu/cgi/viewcontent.cgi?article=1043&context=istpub> [Consultado em 2015-07-04].
- Liu, B. (2010). Sentiment analysis and subjectivity. In Indurkha, N. & Damerau, F., editors, *Handbook of Natural Language Processing*, chapter 26, pp. 627–666. CRC Press, Boca Raton, 2 edition.
- Ljunglöf, P. & Wirén, M. (2010). Syntactic Parsing. In Indurkha, N. & Damerau, F., editors, *Handbook of Natural Language Processing*, chapter 4, pp. 59–91. CRC Press, Boca Raton, 2 edition.

- Mala, V. & Lobiyal, D. K. (2015). Concepts extraction for medical documents using ontology. In *2015 International Conference on Advances in Computer Engineering and Applications (ICACEA)*, pp. 773–777, Ghaziabad. IEEE. doi:10.1109/ICACEA.2015.7164807.
- Manju, A. & Valarmathie, P. (2015). Organizing multimedia big data using semantic based video content extraction technique. In *2015 International Conference on Soft-Computing and Networks Security (ICSNS)*, pp. 1–4, Coimbatore. IEEE. doi:10.1109/ICSNS.2015.7292370.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Press, Cambridge University, 1 edition.
- Martin, J. H. & Jurafsky, D. (2000). Speech and language processing: an introduction to natural language processing, computational linguistics and speech recognition. *International Edition*. Disponível em <http://www.mitpressjournals.org/doi/pdf/10.1162/089120100750105975> [Consultado em 2014-12-19].
- Maybury, M. T. (2008). State-of-the-art tools for more efficient information discovery and analysis. *Society for Competitive Intelligence Professionals Annual Conference*, page 8. Disponível em <http://www.mitre.org/publications/technical-papers/stateoftheart-tools-for-more-efficient-information-discovery-and-analysis> [Consultado em 2013-05-25].
- Mishra, S. & Jain, S. (2015). A Study of Various Approaches and Tools on Ontology. In *2015 IEEE International Conference on Computational Intelligence Communication Technology*, pp. 57–61. IEEE. doi:10.1109/CICT.2015.43.
- Mittal, S., Joshi, K. P., Pearce, C., & Joshi, A. (2016). Automatic Extraction of Metrics from SLAs for Cloud Service Management. In *2016 IEEE International Conference on Cloud Engineering (IC2E 2016)*, pp. 1–4, Berlin.
- Moen, H., Peltonen, L.-M., Heimonen, J., Airola, A., Pahikkala, T., Salakoski, T., & Salanterä, S. (2016). Comparison of automatic summarisation methods for clinical free text notes. *Artificial Intelligence in Medicine*, 67:25–37. doi:10.1016/j.artmed.2016.01.003.
- Mukhopadhyay, D. & Shikalgar, S. (2013). A Model Approach to Build Basic Ontology. *CoRR*, abs/1311.6. Disponível em <http://arxiv.org/abs/1311.6245> [Consultado em 2015-04-17].
- Nhon Do & LongVan Ho (2015). Domain-specific keyphrase extraction and near-duplicate article detection based on ontology. In *The 2015 IEEE RIVF International Conference on Computing & Communication Technologies - Research, Innovation, and Vision for Future (RIVF)*, pp. 123–126, Can Tho. IEEE. doi:10.1109/RIVF.2015.7049886.
- Noy, N. F. & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology. Technical report, Stanford University. Disponível em <http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf> [Consultado em 2015-07-05].
- Olival, F., Garcia, L. D., Lopes, B., & Sequeira, O. (2013). Testemunhar e ser testemunha em processos de habilitação (Portugal, século XVIII). In *Honra e sociedade no mundo ibérico e ultramarino: Inquisição e Ordens Militares (séculos XVI-XIX)*, pp. 315–349. Caleidoscópio, Casal de Cambra. Disponível em <http://hdl.handle.net/10174/9570> [Consultado em 2014-12-19].
- Palmer, D. (2010). Text Preprocessing. In Indurkha, N. & Damerau, F., editors, *Handbook of Natural Language Processing*, chapter 2, pp. 9–30. CRC Press, Boca Raton, 2 edition.
- Patil, A., Pharande, K., Nale, D., & Agrawal, R. (2015). Automatic text summarization. *International Journal of Computer Applications*, 109(17):18–19.
- Pérez, J., Arenas, M., & Gutierrez, C. (2006). Semantics and Complexity of SPARQL. In *International semantic web conference*, pp. 30–43. Springer, Springer Berlin Heidelberg. doi:10.1007/11926078\_3.



- Pintus, R., Yang, Y., & Rushmeier, H. (2015). ATHENA: Automatic Text Height Extraction for the Analysis of Text Lines in Old Handwritten Manuscripts. *Journal on Computing and Cultural Heritage (JOCCH)*, 8(1):1–25. doi:10.1145/2659020.
- Platzer, A. (2013). Lecture Notes on Lexical Analysis. Disponível em <https://www.cs.cmu.edu/~fp/courses/15411-f13/lectures/07-lex.pdf> [Consultado em 2016-04-13].
- Porfírio, L. & Bidarra, J. (2006). Extração de Informação como um Importante Recurso para a Interpretação de Textos. *Revista Línguas e Letras*, 7(12):139–158. Disponível em <http://e-revista.unioeste.br/index.php/linguaseletras/article/view/895> [Consultado em 2015-07-05].
- PostgreSQL Global Development Group (2014). PostgreSQL. Disponível em <https://www.postgresql.org/> [Consultado em 2015-12-15].
- Pretorius, A. J. (2004). Ontologies-Introduction and Overview. *adapted from: Pretorius, AJ, Lexon Visualisation: Visualising Binary Fact Types in Ontology Bases, MSc Thesis, Vrije Universiteit Brussel*, pp. 1–13. Disponível em [http://www.starlab.vub.ac.be/teaching/Ontologies\\_Intr\\_Overv.pdf](http://www.starlab.vub.ac.be/teaching/Ontologies_Intr_Overv.pdf) [Consultado em 2015-07-04].
- Quaresma, P., Rodrigues, I., Prolo, C., & Vieira, R. (2006). Um sistema de Pergunta-Resposta para uma base de Documentos. *Letras de Hoje*, 41(2):43–63. Disponível em [https://www.academia.edu/2674768/Um\\_sistema\\_de\\_Pergunta-Resposta\\_para\\_uma\\_base\\_de\\_Documentos](https://www.academia.edu/2674768/Um_sistema_de_Pergunta-Resposta_para_uma_base_de_Documentos) [Consultado em 2013-05-24].
- Ramkumar, A. S. & Poorna, B. (2015). Semantic Information Retrieval Based on Domain Ontology. *International Journal of Computing Algorithm*, 4(March 2015):1387 – 1390. Disponível em <http://www.ijcoa.com/papers/vol4march/paper55.pdf> [Consultado em 2015-07-05].
- Ranchhod, E. (2001). O uso de dicionários e de autómatos finitos na representação lexical das línguas naturais. *Tratamento das Línguas por Computador. Uma introdução à Linguística Computacional e suas aplicações*, pp. 13–47. Disponível em [http://label2.ist.utl.pt/emr/aulas/lc/emr\\_2001.PDF](http://label2.ist.utl.pt/emr/aulas/lc/emr_2001.PDF) [Consultado em 2014-12-19].
- Reshamwala, A., Mishra, D., & Pawar, P. (2013). Review on natural language processing. *IRACST Engineering Science and Technology: An International Journal (ESTIJ)*, 3(1):113–116.
- Savoy, J. & Gaussier, E. (2010). Information Retrieval. In Indurkha, N. & Damerau, F., editors, *Handbook of Natural Language Processing*, chapter 19, pp. 455–484. CRC Press, Boca Raton, 2 edition.
- Shah, P., Kongari, B., & Dedhia, R. (2015). An Approach to Semantic Search Engine. *International Journal for Advance Research in Engineering and Technology*, 3(3):15–17. Disponível em <http://www.ijaret.org/3.3/AnApproachtoSemanticSearchEngine.pdf> [Consultado em 2015-07-05].
- Stanford Center for Biomedical Informatics Research (2014). Protégé. Disponível em <http://protege.stanford.edu/> [Consultado em 2014-12-17].
- Tawfik, A., Giunchiglia, F., & Maltese, V. (2014). A Collaborative Platform for Multilingual Ontology Development. *International Journal of Social, Education, Economics and Management Engineering*, 8(12):3747–3756. doi:10.1145/1041410.1041420.
- Tosi, D. & Morasca, S. (2015). Supporting the semi-automatic semantic annotation of web services: A systematic literature review. *Information and Software Technology*, 61:16–32. doi:10.1016/j.infsof.2015.01.007.
- Tufis, D. (2009). Algorithms and Data Design Issues for Basic NLP Tools. *Language Engineering for Lesser-studied Languages*, 21:3–50. doi:10.3233/978-1-58603-954-7-3.



- Uschold, M. & Gruninger, M. (2004). Ontologies and Semantics for Seamless Connectivity. *ACM SIGMOD Record*, 33(4):58–64. doi:10.1145/1041410.1041420.
- Valaski, J., Malucelli, A., & Reinehr, S. (2012). Ontologies application in organizational learning: A literature review. *Expert Systems with Applications*, 39(8):7555–7561. doi:10.1016/j.eswa.2012.01.075.
- Van Heijst, G., Schreiber, A., & Wielinga, B. J. (1997). Using explicit ontologies in KBS development. *International Journal of Human-Computer Studies*, 46(2-3):183–292. doi:10.1006/ijhc.1996.0090.
- Viswanathan, V., Rajani, N. F., Bentor, Y., & Mooney, R. (2015). Stacked Ensembles of Information Extractors for Knowledge-Base Population. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL-15)*, pp. 177–187, Beijing, China.
- W3C (2005). Document Object Model (DOM). Disponível em <https://www.w3.org/DOM> [Consultado em 2016-07-15].
- W3C (2008). SPARQL Query Language for RDF. Disponível em <https://www.w3.org/TR/rdf-sparql-query> [Consultado em 2016-07-15].
- W3C (2013). Web Ontology Language (OWL). Disponível em <https://www.w3.org/2001/sw/wiki/OWL> [Consultado em 2016-07-15].
- W3C (2014). Resource Description Framework (RDF). Disponível em <https://www.w3.org/RDF> [Consultado em 2016-07-15].
- W3C (n.d.). Standards. Disponível em <http://www.w3.org/standards> [Consultado em 2016-07-15].
- Wang, W. & Stewart, K. (2015). Spatiotemporal and semantic information extraction from Web news reports about natural hazards. *Computers, Environment and Urban Systems*, 50:30–40. doi:10.1016/j.compenvurbsys.2014.11.001.
- Wiebe, J. & Riloff, E. (2011). Finding mutual benefit between subjectivity analysis and information extraction. *IEEE Transactions on Affective Computing*, 2(4):175–191. doi:10.1109/T-AFFC.2011.19.
- Wu, S., Liu, J., & Fan, J. (2015). Automatic Web Content Extraction by Combination of Learning and Grouping. In *Proceedings of the 24th International Conference on World Wide Web, WWW '15*, pp. 1264–1274, New York, NY, USA. ACM. doi:10.1145/2736277.2741659.
- Xu, R. & Wang, Q. (2015). Large-scale automatic extraction of side effects associated with targeted anticancer drugs from full-text oncological articles. *Journal of Biomedical Informatics*, 55:64–72. doi:10.1016/j.jbi.2015.03.009.
- Xue, X. & Jiang, L. (2015). Optimizing Ontology Alignments through NSGA-II Using an Aggregation Strategy and a Mapping Extraction Approach. In *2015 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, pp. 349–352, Adelaide, SA. IEEE. doi:10.1109/IIH-MSP.2015.100.
- Zhong, B. T., Ding, L. Y., Love, P. E. D., & Luo, H. B. (2015). An Ontological Approach for Technical Plan Definition and Verification in Construction. *Automation in Construction*, 55(2015):47–57. doi:10.1016/j.autcon.2015.02.002.





UNIVERSIDADE DE ÉVORA  
INSTITUTO DE INVESTIGAÇÃO  
E FORMAÇÃO AVANÇADA

**Contactos:**

Universidade de Évora

**Instituto de Investigação e Formação Avançada — IIFA**

Palácio do Vimioso | Largo Marquês de Marialva, Apart. 94

7002 - 554 Évora | Portugal

Tel: (+351) 266 706 581

Fax: (+351) 266 744 677

email: [iifa@uevora.pt](mailto:iifa@uevora.pt)